Acquisition of Mandarin Chinese Pronunciation by Foreign Learners: The Role of Memory in Learning and Teaching

A Thesis

Presented in Partial Fulfillment of the Requirements
for the degree of Master of Arts in the
Graduate School of the Ohio State University

by
Olga Petrovna Tutatchikova, B.A.

* * * * *

The Ohio State University

1995

Master's Examination Committee:

Galal Walker
Mari Noda
Marjorie Chan

Approved by

Advisor
Department of East Asian
Languages and Literatures

To my parents

ACKNOWLEDGMENTS

I express sincere appreciation to Dr. Galal Walker for his guidance and insight throughout the work. I acknowledge his advising as challenging and encouraging. I am also thankful to Dr. Marjorie Chan who expressed her interest in my research and for her thoughtful comments and suggestions. I offer my sincere thanks to Dr. Mari Noda, who being my instructor in two classes on second language pedagogy stimulated my interest in the field of the psychology of learning and mastering teaching approaches and methods. Thanks go to Dr. Keith Johnson, Dept. of Linguistics, who advised me on the literature in the field of the acquisition of L2 phonology. Gratitude is expressed to graduate students of the Department of East Asian Languages and Literatures: He Dajiang, Wang Jing, Zhang Aiping, Wu Suemei and Jin Shunde, who were very helpful in collecting the data base of most common mispronunciation errors made by foreign learners of Mandarin Chinese and also gave various valuable suggestions for my research. I offer sincere thanks to my friend Chris Lacy who spent much time editing my English and correcting my style. I am thankful to Sergei Klimin, M.A. in Physics and Economics, for his guidance with the Distributed Model of Human Learning and Memory. I offer my thanks to Diane Dagefoerde and Elizabeth Stephan-Conrads in the Humanities computer lab for their big help. To my husband, Sergei Kulyov, I offer sincere thanks for his unshakable faith in me and for his willingness to provide me with the peace and comfort necessary for work. To my son, Arcasha many thanks for his understanding and love.

VITA

February 4, 1963 ....Born -- Gusev, Russian Federation, USSR

1986...B. A., Leningrad State University, Russian
....Federation, USSR

1987-1993 ....Correspondent, Chinese Division, Moscow
....Radio World Service, Moscow, USSR

1991-1993 ....Coordinator, YMCA USA, International
....Office for Europe

1993-Present ....Graduate Student at the Department of East
....Asian Languages and Literatures, The Ohio
....State University

FIELDS OF STUDY

Major Field: Chinese Language Pedagogy

Studies in Chinese Language Pedagogy: Dr. Galal Walker, Dr. Mari Noda
Studies in Chinese Language: Dr. Marjorie Chan, Dr. James Tai

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE | PAGE

Chapter I. Introduction: Towards a model of learning L2 pronunciation

1.1. Introduction

The intent of this Thesis is to investigate the role of memory in learning pronunciation of L2 and, in particular, the pronunciation of Mandarin Chinese. Knowledge of how sounds of L2 are perceived, stored and retrieved by adult learners is important for understanding how the process of learning operates. An investigation of limitations and restraints on learning, as well as discovering the potentials of human memory, is the focus of this research, aiming at bringing updated views and models of memory and learning to foreign language pedagogy.

In this chapter the problem of choosing a balanced approach for acquisition of L2 phonology will be discussed. Two different methods of teaching L2 pronunciation presently in practice, as well as the theories underlying them, will be introduced. We will also present our view on the problem of learning L2 phonology and introduce the theory of human memory and learning that we rely upon.

1.2. Nature and history of the problem

Hector Hammerly (1991: 42) considers that the approach to teaching L2 which focuses on communicative language proficiency removed the concept of linguistic accuracy from a central role in L2 study. According to Hammerly, this approach to learning L2 stresses survival in specified communicative situations in the target culture rather than focusing on accuracy (1991: 41). The ability to function in the target language environment and to fulfill specified goals is considered the ultimate goal of a learner. Thus, the principle of achieving *functionality* built in the idea of considerable inaccuracy in L2 acquisition. This principle can be found in the descriptions of proficiency ratings given by the US Foreign Service Institute (FSI) / Interagency Language Roundtable (ILR) and the ACTFL guidelines.

The approach of focusing on achieving "proficiency" rather than accuracy in L2 phonology acquisition most probably relied upon the traditional strong belief among teachers and learners of a second language (L2) that adult students (of twelve and above years old) can acquire the pronunciation of the second language only to a certain degree of accuracy and cannot resist the interference of the first language pronunciation, i.e., cannot prevent speaking L2 with an accent. This belief came from and was supported by the critical period hypothesis (CPH), which was brought from the field of biology to the field of language learning by Penfield and Roberts (1959).

According to this hypothesis, there is a critical period that terminates about nine to twelve years of age, after which complete or native-like mastery of languages, first or second is difficult or unlikely. They argued that this critical period coincides with a period of neural plasticity - that is, a period during which different areas of brain are able to assume a variety of functions, including language. After neural plasticity is outgrown, the functions of the different parts of the brain cannot be rearranged. The hypothesis was developed further by Lenneberg in *Biological Foundations of Language* (1967). In his book Lenneberg proposed that the critical period for language learning extends from two years of age until the age of puberty. Lenneberg argued that language learning was difficult after puberty because lateralization of language functions in the left hemisphere was thought to be completed by this age. Lenneberg supported this argument with observations of children whose left hemisphere was damaged before age of twelve and whose language functions had transferred to the right hemisphere. Such transfer occurred less frequently in individuals who suffered brain damage after twelve years of age.

Although the critical period hypothesis was supported from the position of general neurological plasticity (Penfield) and from the position of hemispheric function specialization (Lenneberg), the hypothesis was found to be only partially sufficient in answering the questions being raised in the field of L2 acquisition from conceptual and empirical points of view. Among important conceptual weaknesses was the fact that Penfield and Lenneberg relied upon L1 competence which does not necessarily apply to L2 learning. Their observations pertained to adults with *impaired* first language skills as a result of brain damage or pathology. It will not be necessarily true that healthy adults with intact neurological systems would have analogous learning problems. It was proved empirically to be impossible to define exactly when complete lateralization of language function in the left hemisphere is achieved, e.g., the termination of critical period for language learning: Evidence was found that left-hemispheric specialization for language may occur by age five (Krashen 1974) and even may be present at birth (Molfese and Molfese 1979). The assumption that localization occurs at the same time and rate for all subsystems in the brain was also found inconsistent. It was found more likely that the neural plasticity on which the CPH was presumed to change *progressively* with age and may extend beyond puberty and that it is more appropriate to talk about a *sensitive* rather than a *critical* period for L2 learning (Lamendella 1977; Seliger 1978). According to this notion, certain language skills are acquired more easily at particular times of development than at other times and some language skills may be learned after the critical period, although less easily.

A more recent discussion held by Walsh and Diller (1981) about the maturational rates of different types of neurons in the brain proposed an alternative view on neural plasticity. They distinguished two types of neurons in the brain: The first type are neurons that are fully mature and functional early in development -- macroneurons. The other type of neurons develop much more slowly and continue to develop into adulthood --local-circuit neurons. Owing to the slow development of these neurons, neuron plasticity provides for the development of language ability into the adult years.

From the empirical evidence, the CPH theory was also found to be inconsistent in several aspects, although neither of the findings were enough to reject it. The CPH identified two constraints on second language learning. Second language capabilities are more difficult to achieve and require more effort to learn with increasing age. Native-like proficiency, especially in phonology is rarely achieved in adult age. Evidence, however, indicates that native-like proficiency in a L2 phonology is possible after puberty, if a systematic guidance and sufficient feedback are provided. In an experiment conducted by Neufield and Schneidermann (1980) for testing the acquisition of phonology by adult learners, eighteen hours of systematic intensive instruction in pronunciation of Japanese, Chinese or Eskimo were provided to twenty adult English-speaking learners. The learners were subsequently evaluated by *native* speakers of these languages who using their intuition judged that fifty percent of the subjects had native-like pronunciation. In contrast, the results of L2 phonology acquisition by unguided learners of L2 who lived in the target language environment for extended period of time, their pronunciation performances were not reported as accurate, especially in sound production (Flege 1992: 588-589). This suggests an interesting experiment where learners of L2 phonology who experienced systematic guidance during a short period would be compared with learners who lived in the target language environment for a significant time (five years plus) but did not experience any systematic guidance. My observation of a seventeen year old American learner of Chinese in the Department of East Asian Languages and Literatures of The Ohio State University, is suggestive: J.W. lived in Taipei, Taiwan for more than ten years, was formally instructed in English in the local American School. However, he learned spoken Chinese through extensive interaction in the target language environment and by communicating with numerous Chinese friends and neighbors. His pronunciation of Mandarin Chinese was characterized by a strong L1 interference due to substituting L2 phonetic categories with L1 phonetic categories. Furthermore, J.W. had a very vague idea about Chinese tones. His pronunciation of tones and segments could only be graded as below satisfactory and so he had to start learning pronunciation from the beginning level. In contrast to his pronunciation, J.W.'s vocabulary was very rich and could easily be compared to that of native speakers. By the end of the first year of Chinese systematic instruction, J.W. showed a considerable improvement in production skills, as evaluated by native speakers and instructors. The results of experimenters (Neufield and Schneidermann 1980; Flege 1992) and my observation of learners of Chinese lead me to the following hypothesis: Acquisition of accurate L2 pronunciation is possible in adults. Passing over the age of puberty does not imply an immediate loss of the ability to acquire behavior that approximates the phonology of L2 speakers. Pedagogical instruction is crucially important to achieving accuracy in an adult's learning of L2 pronunciation.

Observation of learners' behavior as well as new findings in the field of neuropsychology concerning neuron plasticity in adults and its role in L2 learning generally support a different approach in L2 pronunciation learning contrasting with the method of achieving "proficiency" or "functionality" discussed above. This method was called "Step-By-Step Linear Learning method" of an L2 acquisition by Hammerly (1991: 41). The method implied that L2 phonology can be acquired in adult age if a thorough bottom-up learning of segments and tones is trained. Being a considerably slower than proficiency method it requires acquisition a solid theoretical background for pronunciation as well as employs a lot of drills. On the other hand, this method was criticized by Hammerly for not paying much attention to developing spoken interaction and communication skills. As Hammerly described his observations of this method, graduates of this method could correctly produce segments, but could not communicate in the language (Hammerly: 30).

1.3. Problem statement

The new findings in the areas of neuropsychology and neurolinguistics focused on language learning and memory scientifically support the belief that the age of puberty is not a deadline after which acquisition of accurate native-like pronunciation is impossible. The problem appears to be not in the physiological fossilization of the articulation system after puberty, as it was considered earlier, but in most part is in the availability of the knowledge of how L2 phonology is perceived and stored in the memory and how teaching methods can not effectively impose on the process of learning in adults. The development of the method of learning and teaching pronunciation must be based on updated theories of memory and learning in humans. Although the investigations in the fields of memory and learning cannot give the ready answers about the design of teaching methods they can challenge our views and beliefs about learning and teaching L2 phonology and stimulate research in this area.

1.4. Statement of Purpose

The purpose of the thesis is to propose a model of learning L2 pronunciation. For designing this model we have adopted the concept of memory and language learning designed as the result of experiments, observations and generalizations. In order to verify the applicability of our model of learning and memory in practical pedagogy, we analyzed how well this model fits into the present curriculum of teaching Chinese in The Ohio State University. On the basis of the Intensive Chinese Curriculum Design created by Galal Walker (1989) and relying upon observed patterns of mispronunciation at different levels of Chinese learning, we proposed the guidelines for the four levels of Chinese instruction which could be implemented into the program developed for the Curriculum Design.

Chapter II. Information processing and learning in human memory

2.1. From the history of the study of memory: Localizationist and distributed views.

Two contrasting ideas can be identified in the history of the memory and brain sciences (Squires 1987:56-75), which have been concerned with the location and the type of memory store. One tradition is identified as *localizationist* or *deterministic* which claimed that the nervous system is composed of identifiable, localized parts and that behavioral functions can be localized to particular components of the nervous system. This school of thought achieved considerable success by demonstrating that localized brain lesions or stimulation produce highly specific effects on different patterns of behavior such as language, vision or motor movement. The other tradition, a *probabilistic* or *distributed* view of memory, has developed an opposite opinion on how memory is stored in the brain. This theory has argued that behavior and mental activity arise from the integrated activity of the entire brain. This theory's fundamental concept was that memory must be widely and equivalently distributed throughout many areas of the brain and that homogeneous neurons and synapses participate in storage and recall processes. Recent experiments have found that both theories are compatible in many ways. First of all, the brain was experientially proved to be divided into numerous specialized functional areas. For example, it was demonstrated by the lesion data from aphasic patients that the representation of acquired language is organized in the specialized areas of the left cerebral hemisphere. Another observation supporting the localizationist view was that localized stimulation in bilingual patients can cause naming errors in one language but not in another. Frequently only one of several language functions, e.g., naming, reading or phonetic identification, is altered by stimulation at a particular site. On the other hand, it is impossible to deny that memory representations are distributed across a large collection of common neurons. For example, anomia caused by stimulating one particular area of brain, produced difficulty in remembering the names of things in one language, but not in one isolated group of names in the language.

The conflicting views on storing information in memory can be considered reconcilable at the present level of the knowledge of memory. According to the recent findings, the size of the functional areas across which memory is equivalently distributed and how extensive a class of information can be equivalently represented in the same set of neural elements need to be closely investigated in each concrete case. At the present time, the generally accepted view of memory is one in which memory is widely distributed in the brain, but different loci store different aspects of the whole. The principle of distributed memory proposes, "that items in memory are represented by different patterns of activity among the *same* neural elements rather than by occupying different loci that can be linked together to form associations" (Squire: 67).

The distributed model of human learning and memory, proposed by J.L. McCelland and D.E. Rumelhart (1986:170-215), is now one of the popular models of memory and learning among the existing models of distributed human memory (Squire: 67). Its main contribution was the development of the principle of interactive activation of modules and the change in the weights of the interconnections of units activated by external input as the basis of input processing and learning. The process of learning was described by the delta rule, which explained the process of learning as the general ability of a network of interconnections to adjust to external input by decreasing the discrepancy between desired output and actual output. The effect comes from two sources: first, from external input and second, from internal input from other modules activated by the external input. The result of the process is the strengthening of the interconnections across the units which were activated by the external stimuli.

Although the model of learning and memory was developed as a psychological model and not for application in pedagogy, it explains important issues about learning processes in the memory of a learner which can help us take a new look at pedagogical approaches and methods. The subject of both disciplines is common -- the development of human memory. The detailed assumptions of the model give us an opportunity to test existing pedagogical approaches in teaching Chinese pronunciation and to investigate new possibilities in introducing material and developing new skills in learners.

2.2.1. The Distributed Model of Human Learning and Memory: How the information is stored and retrieved

The distributed memory system is presented as a microstructure of an assumed model of memory. The model is a network of highly interconnected units. The units are organized into modules which communicate with each other and with units in other modules. Each module represents a synthesis of the states of all the modules it receives inputs from. Modules combine inputs from a number of different sources of information: from outside (external input), and from inside (from other modules and units). The processing of input is measured by the change in the weights of interconnections among units.

During processing cycle, each unit determines its net input on the basis of external input and the activations of all other units involved in the process at that moment. The net input will be modulated by the summarized weights and will determine the strength and direction of each unit's effect on every other unit. The activations of all the units are processed simultaneously. After a particular unit is activated by input, its activation level will change correspondingly to the inputs coming from other units and modules, until the weights of interconnections are adjusted.

The delta rule of learning determines several important factors that affect memory processing and learning: First of all the input is divided into internal and external. Internal input is the sum of different inputs arising from the other units in the module. The value of the summarized input is determined by the independent values of the internal and external inputs. This value causes the certain level of activation which leaves traces in memory. These memory traces are described as changes in the entire set of weights in the module. The strength of memory traces are proportional to the adjustments of weights in the interconnections across units.

According to this view, no single unit is a carrier of some particular pattern of information, such as a word, a visual image, a sound, but each input pattern is decoded into microcomponents and then distributed among corresponding units, which equally store different information patterns. The original form of the input is retained in the interconnections of different units. For example, if a new word was initially perceived in a visual form, then the interconnections among the units corresponding to the visual modality would be stronger than if this word would have been perceived auditorily first and then visually. This will be true at least until further inputs of the same word strengthen other interconnections corresponding to this pattern (meaning, auditory, etc.). After being decoded into microcomponents and located in the corresponding units in memory, each representation is completely or partially restorable in its original form. The input stimulus "fires" the appropriate network of interconnections among the units that took part in decoding the information into microcomponents. This structure explains why, when recalling a word, we not only remember its meaning and verbal representation, but very often recall such additional information as the situation in which the word was first heard, the context and even the reasons why this word struck our attention at that moment. The reason a word can associate so much "redundant" information around itself may be the lack of input weight in terms of its meaning or verbal form at the first acquaintance with the word, which is essential for its successful processing and retention. For example, if a word's meaning was not sufficiently clear from the context, then the interconnections within the units corresponding to the semantic memory would not be strong enough for a future retrieval of the word. In this case, memory demands more contextual information about the new word which would enable the network of interconnections to abstract the central tendency of its application to make the interconnections stronger. Memory *compensates* for the gap in the input information about a word by providing additional data in order to stimulate completion of processing and, as the ultimate goal, the word's retrieval.

Another example that can illustrate how memory compensates for the lack of some kind of external input weight by retaining and retrieving supplemental information is the "tip-of-the-tongue" phenomena described by Donahoe and Wessels (1980: 341). This phenomena, based on the feeling that we know a particular word or name, but cannot remember it, was examined by giving adults dictionary definitions for uncommon words. After a while the subjects were asked to recall the words that were defined for them. If subjects said that they were certain they knew the word but could not pronounce it, then those subjects were said to exhibit the tip-of-the-tongue state. At the same time the subjects were able to indicate which words were verbally similar or related in meaning to the word they were trying to remember; they could even point the location of the word in the list they learned it from. According to these observations, Donahoe and Wessels suggested that a word is represented in memory as a group of features that may be phonological, visual or semantic in nature. This group of features represents the useful information for strengthening the input pattern that can be easier to retrieve in the future.

Retrieval is the process of the re-creation of a prior pattern of activation stimulated by external input (McClelland and Rumelhart 1986: 178). The external input, in order to cause retrieval, must use a fragment of the original state as a "retrieval cue". The external input is called *excitatory* input. Its role in the process of retrieval is to influence the internal connections in such a way that when a part of the state is presented, the internal connections will tend to reproduce the rest. When the excitatory input pattern reaches the corresponding network of units, they take on activation values appropriate for the particular pattern. The units recreate the external input by adjusting the weights in the internal connections (internal input) to the external input. The process of adjusting the weights depends upon the difference between the external and the internal inputs. This implies that the process of recreating (or decoding) information goes in the opposite direction from that of encoding it: If for determining its activation the unit must be able to aggregate both internal and external input, then for adjusting its weights it must be able to distinguish between external and internal inputs. The delta rule can determine the size and direction of the internal input necessary for the retrieval process.

If the weight adjustment is δ_i, the difference between the external input to the unit and the net internal input to the unit from other units in the module will be:

$$\delta_i = e_i - i_i$$

where e_i is external input and i_i is internal input. If the difference is positive, then the internal input is not activating the unit sufficiently and the retrieval would not be complete. If negative, it is activating the unit too much. If it is zero, everything is fine and the retrieval will present representation of good quality. The magnitude δ_i

determines the size and direction of the overall change that needs to be made in the internal input to the unit. To achieve this overall effect, the individual weights are then adjusted according to the delta rule formula (McClelland and Rumelhart: 180):

$$\Delta w_{ij} = \eta \delta_i a_j$$

The parameter η is a global strength parameter which regulates the overall magnitude of the adjustments of the weights; The parameter Δw_{ij} is the change in the weight to i from j. Thus the delta rule tends to drive the weights to the right values to make the internal input to a unit match the external input. The process of learning is explained by the ability of a network to adjust the weights of interconnections in the module so that internal input will tend to reproduce the external input to that unit.

The retrieval mechanism, described by the delta rule is limited to the *linear predictability constraint*. The constraint implies the following: " Over the entire set of units the external input to each unit must be predictable from a linear combination of the activations of every other unit" (McClelland and Rumelhart: 181). However, it was acknowledged by the authors of the model that this is the limitation of the mathematical description of the model of learning, rather than the human learning itself. The authors distinguished between the *described* model and the *human* cognitive system which has broader learning opportunities and cannot be entirely limited to this constraint.

The linear predictability constraint was taken as a basis for his step-by-step method of introduction of new information proposed by Hammerly (1991:142). He pointed that the amount of information about the second language introduced to students must grow successively. Starting with very basic and widely used structures, students must first gain complete control of them through the stage of manipulation on the basis of the introduced patterns. Hammerly brought up examples from his teaching experience which indicated that nothing is learned well if there is a hurry in introducing material without a sufficient review and feedback. Students should learn each rule and element thoroughly the first time they are introduced to it, building up and *integrating* knowledge into the structure of previously learned knowledge. Hammerly proposed a general cycle of learning which ideally fits the linear predictability constraint. According to this cycle, each rule must go through the three main stages of learning: The first step in the proposed cycle is material presentation, which consists of exemplification and achieving understanding of the material by students. The second step is learning of the presented material through mechanical drills and meaningful exercises. Interactive application is the third step: First in communicative activities and then to "real" communication. This strategy clearly follows the principle of gradual building material in a student's memory, confirming the assumption that when the external input is predicted from a linear combination of activations of units, then it can be perfectly learned.

2.2.2. The process of decay in the Distributed Model of Human Learning and Memory

Each trace in memory undergoes a process of decay (McClelland and Rumelhart: 181). New traces first decay rapidly, but then the remaining portion becomes more and more resistant to further decay. At the present state of knowledge it is still impossible to say if it ever reaches the point where there is no decay at all. The authors of the distributed model of human learning and memory assumed that each input exerts large short-term effects on the weights, but after these effects decay, the residual effect is considerably smaller. Each increment has its own temporal history which depends on many factors. The amount of internal and external inputs into the particular module interconnections is most important in predicting the strength of memory traces and decay rate.

2.3. Information processing and human behavior

The process of storing information in the human brain discussed above has also been widely investigated in human behavior through observations and experiments. Cognitive scientists and psychologists are seeking ways to detect memory processes through testing different aspects of human behavior in response to various stimuli. As experiments showed, storing information is closely associated with a particular modality in memory, or to the processing systems used to analyze the incoming information (Sperling 1967). Auditory modality in memory is used, for example, for recognizing people in another room or over the phone by their voices or by identifying other people's origins by their dialects and accents. The following section focuses on processing auditory information.

2.3.1. Stages of processing auditory information: Sensory store and echoic memory

Auditory information is perceived and initially stored in the sensory store of memory corresponding to audition (Loftus and Loftus: 11). The part of the sensory store corresponding to audition is called echoic store or echoic memory. The early investigation of this step in processing auditory information is associated with Moray, Bates and Barnett (1965) and was extended by Darwin, Turvey and Crousder (1972). By echoic memory they referred to the short lasting trace of recently presented sounds. It was discovered that information remains in sensory store for four seconds before it completely decayed (Darwin et al. 1972). Information cannot remain in the memory longer unless it undergoes a deeper processing and/or is maintained through rehearsing.

In order to enable further processing the auditory information must be recognized as corresponding to some information already existing in memory. Recognition means attaching meaning to the incoming physical pattern. The phenomena of pattern recognition was closely discussed by Lindsay and Norman (1972). They found that after the sensory input was "pattern recognized" as a meaningful unit, it would be retained in memory for a longer period. The stage of processing the input information is called short term store or short term memory (STM). The following example from a pre-school practice can illustrate the process of pattern recognition: When a teacher introduces a letter to children, for example, the letter "t", in order to facilitate the memorizing process the teacher will name some familiar words starting with the letter "t": table, telephone, teacher, time. When associated with certain meanings, the letter will be kept longer in memory. Loftus and Loftus (1976: 27) proposed a model of pattern recognition which explained the pattern recognition process as a medium between echoic store and STM. According to this model for recognizing a pattern, the program must ask a series of questions which will enable it to map the auditory input completely or partially onto previously stored information.

Sperling et al. (1971) (summarized by Loftus and Loftus: 27) reported results that give a clue to understanding how long it takes to process one pattern of information and how many patterns in a row can be processed and stored in STM. In his experiment, subjects were presented with a series of arrays, four to six letters each. The arrays flashed sequentially on a cathode-ray screen. The catch was that one location of one of the arrays contained not a letter, but a digit. The subject was supposed to report which location contained a digit. The interstimulus interval varied from 5 to 320 milliseconds. The experimenters assumed that for locating the digit in the arrays of letters subjects had to pattern recognize the items in the array. Sperling et al. used the probability that the subject correctly reported the position of the digit in order to estimate how many items had been pattern recognized from each array. For example, if the probability was reported 0.67, it could be inferred that the subject has pattern recognized two-thirds of the items in each array, or six out of the nine items in the array. Figure 1 shows the estimated number of items scanned as a function of the interstimulus interval. There are several interesting characteristics of this data. First, when arrays are presented for 40 milliseconds apiece, the subject reported to recognize about five items from each array. At this interstimulus interval the subject , therefore, scanned about one symbol every 8 milliseconds. The second important result of this experiment is that the curve flattens out in the point of 150 milliseconds of interstimulus interval. Increasing exposure time (interstimulus interval) beyond a certain point does not seem to lead to further pattern recognition of the patterns in the stimulus arrays. In this experiment the subject was not required to remember the symbols he was pattern recognizing, as the task was only to find the digit. Therefore, the short-term memory could not fill up. The results of the experiment suggested that the levelling off of the curve in Figure1 is not caused by a limited capacity of STM, but by how long the pattern recognition program can operate before it needs a rest. Therefore, the alternative theory which claimed that short term memory had a limited capacity was discredited.

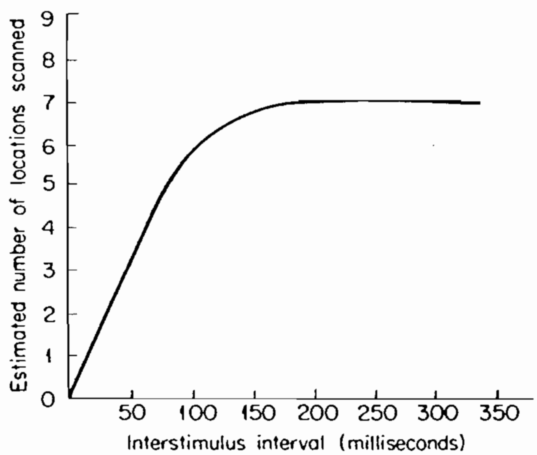

Figure 1. Results of the Sperling et al. (1971) experiment.

2.3.2 How information from sensory store is transferred to short-term memory (STM)

Information in the sensory store can be held for about four seconds before its complete decay. During this time, the useful information must be reinforced and the rest let decay. Loftus and Loftus (1976:30) noticed that a person chooses which information to attend to and to transfer to the STM. The decision is made on the basis of the task a person is trying to accomplish. Loftus and Loftus explained the process of choosing the necessary information from the sensory memory and transmitting it to the short term memory as the result of analyzing information for its meaning.

2.4. Short-term memory

The view of the STM has been changing during the last 25-30 years. There are two main views: one view constituted the STM as a single, phonologically based memory. Its main function was- considered to be the maintaining of auditory information through rehearsal. The other hypothesis considered STM as a collection of temporary working capacities corresponding to particular information processing modalities. STM was considered as a part of unitary memory processing: short-term and long-term memory. Under this view the term "short-term memory" referred to the system that retained information during a limited period while it became incorporated or transferred into a more stable, potentially permanent store (Summarized by Squire: 138). The recent view on STM considers it to be a multicapacity structure, with the capacities intrinsic to information-processing subsystems. The short term storage of heterogeneous information is called working memory and is considered as a part of the human memory working as a "buffer" which maintains information while it is being processed (Baddeley and Hitch 1974). The auditory-verbal STM under this view represents the temporary working memory capacity of just one particular language-based processing system.

2.4.1. Phonological, visual and semantic encoding in STM

. The two views on STM agree on the fact that language based information is represented in verbal form. Since there was no agreement on whether the verbal code is *acoustic* (Conrad 1964) or *articulatory* (Hintzman 1967), the more general term *phonological encoding* is used (Donahoe and Wessels: 443). Evidence for phonological encoding has come primarily from an analysis of the types of errors that subjects made in STM experiments.

Conrad (1964) proved that errors are acoustically related to correct responses in a memory span experiment. In his experiment, he gave the subjects a letter string to remember: A C Q F G H J P. The subjects may substitute a "T" for the "C" and recall "A T Q F G H J P", which is acoustically and articulatory similar to the tested stimuli. It was further discovered that errors tend to be acoustically related to correct responses at short, but not at long, retention intervals (Conrad 1967). At short retention intervals usually only one phoneme is forgotten, so the subject makes a guess on the partial acoustic information and replaces a forgotten phoneme with an acoustically similar one. At long retention intervals, as a rule, many phonemes are forgotten and guesses are random.

Another experiment that proved the phonological encoding of visual information in STM was conducted by Conrad and Hull (1964). In the experiment, two strings of letters were presented to subjects, one of which was acoustically similar (for example, E G C Z D B G) and the other acoustically dissimilar (for example, F G O A Y Q R). Performance on the acoustically similar strings appeared considerably poorer than performance on the acoustically dissimilar strings. The result confirmed the observation that phonological encoding of visual stimuli was the decisive factor for remembering the strings in the case of a short retention interval.

An experiment conducted by Yi Xu (1991) on the material of Chinese characters confirmed the phonological encoding of visual information. He tested the hypothesis that tonal similarity of characters affects STM of verbal material in a tone language. In the first experiment, subjects were visually presented with eight sets of seven monosyllabic Chinese morphemes which had either the same tones (1) or different tones (2):

(1)

班	肝	安	贪	山	翻	三
bān	gān	ān	tān	shān	fān	sān

(2)

肝	残	展	三	慢	然	反
gān	cán	zhǎn	sān	màn	rán	fǎn

More errors were made on the recall of the monotonal sequences than on multitonal sequences, which confirmed the hypothesis of tonal effect on sensitivity of STM. In the second experiment subjects were presented with 48 sequences of disyllabic nonsense words. Each sequence contained three nonsense words. One half of the sets presented the tone sandhi rule in Mandarin Chinese (1) and the other half did not (2):

(1)

常肯	肠者	厂把
chángkěn	chángzhě	chǎngbǎ

(2)

常并	肠日	厂寸
chángbìng	chángrì	chángcùn

Subjects were required to recall the sequences from the STM. The consequence of applying the tone sandhi rule is that all the first characters in the sequences become identical in pronunciation and create potential phonological confusion. More errors were reported on the recall of the sequences subject to the tone sandhi rule than on those not subject to it. Thus the existence of a surface phonetic representation of visual information in STM was confirmed. The results of this experiment confirmed the phonological mechanism of STM.

In addition to *phonological*, STM implies *visual* and *semantic* encoding of incoming information. Visual encoding in STM was tested in an experiment conducted by Cooper and Shepard (1973) who studied the changing of a reaction time of the subjects on the rotation of a stimulus from its covert, upright position.. The results of the experiment showed that the reaction times increased as a function of the difference in orientation between the presented letter and the letter in the normal, upright position. The greater the angular departure of the test letter from the normal, upright position, the longer the reaction times were. This experiment confirmed the view that performance in STM requires a short-term visual code.

Semantic encoding in STM was observed in an experiment conducted by Shulman (1972). In his experiment subjects were presented with a ring of ten spoken words at the rate of two words per second for each of the trials. Immediately following the tenth word in a trial the subject saw "I" or "M" followed by a probe word. "I" indicated that the subject was to report if the probe word was identical to one of those ten words he heard. "M" indicated that the subject had to identify whether the probe word had the same meaning as one of the ten stimulus words. The results revealed that in "I" trials when subjects had to report if the probe word was identical to a stimulus, the subjects would often confuse a synonym with the identical word and say "yes" to the synonym. The semantic similarity of the probe word and the synonym on the list led the subjects to confuse the semantically related words as identical. The semantic encoding in STM was also confirmed by later experiments (Wickens 1972).

The problem of the STM capacity was widely discussed in the sixties and in the beginning of the seventies (Atkinson and Shiffrin 1968; Miller 1968). Atkinson and Shiffrin in their experiments supported the view that retention in STM was strongly related to consciousness. It was considered plausible that the capacity of the STM was bounded by the limitations of consciousness. The human consciousness was considered to be limited in capacity to contain only a certain amount of information patterns (Atkinson and Shiffrin 1968). It was also agreed that the consciousness could retain information only during a limited period of time, after which the information either decays or is transferred to the long-term memory (LTM). Hayes (1952) and Pollack (1953) observed a very interesting fact: Subjects could retain in the short-term memory around seven items, no matter how much information each item contained.

In 1956 George Miller in his study "The magical number seven, plus or minus two", was the first to propose that stored the information in STM is organized in chunks. "Chunk" stands for a unitary representation of a piece of information in memory. A "chunk" may be a letter, a phoneme, a digit, and it may also be a word, an idiom, a proverb, or a song. Everything that is considered as a unit is termed a chunk of information. STM was discovered to be able to hold about seven chunks at a time, but much less of randomly organized groups of information. For example, a subject was asked to remember and report back a string of five groups of letters: FB - ITW - AC - IAIB - M. The subject almost certainly will not be able to remember all the twelve letters divided into five groups in this way. But if the letters will be grouped differently, for example: FBI - TWA - CIA - IBM, the subject will remember all the five groups of letters without much effort. The reason was in the appropriate grouping or chunking of the information which made remembering most effective. (Loftus and Loftus 1976: 47). The phenomenon of the chunking of randomly organized information through attaching meaning to each pattern let it be pattern recognized. Thus we can consider chunking as a device that facilitates remembering through pattern recognition. The hypothesis of chunking information was also tested and confirmed on Chinese character material by Zhang and Simon (1985). In their experiment subjects were introduced to different sequences: radicals, characters and words. The data of the experiment confirmed that students could retain in memory up to seven chunks of phonologically presented information, regardless of what it was: A pronounceable radical, a character or a two-syllable word. However, when chunks of information were not phonologically encoded, (unpronounceable radicals), then subjects could remember only up to three items.

In 1967 George Sperling proposed a model of short-term memory organization. (Figure 2.) Here the part of the model which corresponds to the processing of auditory information in memory, e.g., to the echoic memory, is reproduced.

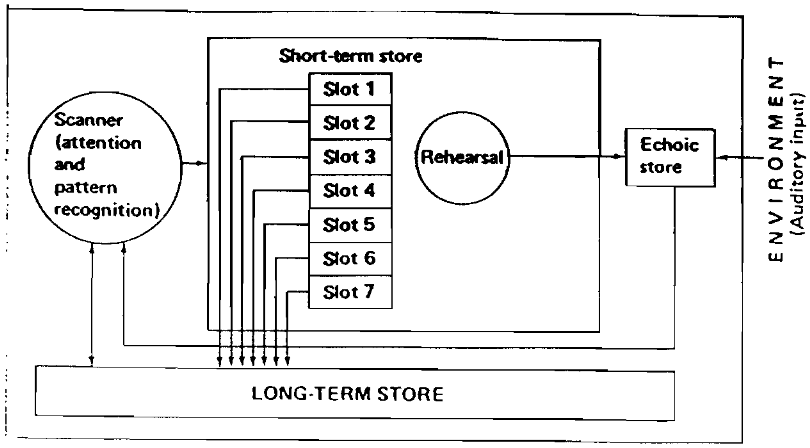

Figure 2. A model of short-term auditory store (Adapted by Loftus and Loftus from Sperling, 1967).

This model incorporated all the characteristics of the STM: rehearsal, scanning, limited capacity and the integration with the long-term memory. As is reflected on the model, the auditory input goes directly into the sensory store, where the information is pattern recognized or "scanned". The circles in the figure mean *processing* information, whereas boxes *holding* information. The result of scanning is filling up the available slots in the STM memory with one chunk of information per slot. The information which fits into slots is not the actual sensory information as in the sensory store, but is a functional pointer to the particular store in the LTM. The arrows from each slot to the long-term memory indicate these connections. The next step of the model is a rehearsal process. The rehearsal process goes successively through each slot creating an auditory representation of the information identified by the slot. The auditory representation then acts as a new auditory stimulus and is placed into echoic store where any auditory stimulus is placed.

The scanner then reoperates on echoic store replacing information into its appropriate slot, and the information in short-term store is thereby continuously recycled.

2.4.2. Retrieving information from STM

As the capacity of the STM is limited, the information can be retrieved rapidly and without much effort. The question that was investigated considered the basic principles of the retrieval process: "Is the desired information searched in parallel and is simply "plucked out", or it is necessary to search sequentially (item by item)?" (Loftus and Loftus: 52). Sternberg (1966) proposed that by looking at the function relating the reaction time to the size of the memory set it was possible to distinguish between the parallel and sequential hypotheses. He conducted an experiment in which he could evaluate the function of the search time to memory set size by varying the amount of information in the short term store. The function showed that an increase in the number of items in the short term memory increased the duration of the search time. Based on this observation, Sternberg drew a conclusion that information in the short-term memory is searched sequentially. If to assume that the information was searched in parallel order, then the increasing data in the STM would not affect the time of search. Another observation about this function allowed Sternberg to conclude that the search is *exhaustive* in character, as opposed to the other view that considered the search *self-terminating*. The function reflected that even when a match between the control item and one of the memory set items was found, the subject continued to compare the control item to the rest of the items. It showed that the subject did not stop the search after the match was found, which would seem reasonable, but continued comparing. Sternberg explained this search mechanism by assuming that it may be easier to wait until all the comparisons have been made and then to make a decision, rather than to make a decision after each comparison. Thus, the retrieval process from the STM was characterized by Sternberg as a sequential, exhaustive search.

2.5. Long-term memory (LTM)

There is no unified way of characterizing long-term memory. The traditional way of commenting on the long-term memory is to view it as a human ability to recall in order to interact with the environment in a dynamic way.

2.5.1 The information types in the LTM

According to the model of memory proposed by Sperling (1967) and described above, STM is interconnected with LTM. The first time LTM is involved in information processing is during "scanning". Scanning occurs when the input is pattern recognized in terms of its meaning, or in terms of sensory familiarity with the information previously stored in LTM. The choice of which of the particular patterns of incoming information needs to be processed further is made on the basis of scanning. LTM stores informationat least two ways: semantic and sensory (phonological and visual).

A number of experiments have suggested that the dominant mode of representation in long-term memory is semantic (Anisfeld and Knapp 1968; Baddeley 1966; Grossman and Eagle 1970). The experiments used the method of false recognition, under which the kinds of errors made by subjects were examined. The experiments used a long retention interval as a part of the procedure. In one of the experiments subjects saw the word "car" as a prime and then saw two ambiguous related control stimulus - "auto" or "tar". Recognition was easier and reaction time was shorter in the case when the "auto" control stimulus was presented. This facilitation was explained because the two words, "car" and "auto" were semantically related. The "tar" control, which was phonologically similar, did not show the same level of facilitation.

Another type of experiment was conducted by Kintsch and Buschke (1969). Subjects had to remember a list of words and report them back in the same order they found them. Kintsch and Buschke showed in this experiment that semantically related words were harder to remember in the correct order. They argued that the words in one category, such as "car" and "auto", when recalled after a long retention interval (recalled from long-term memory), would interfere with each other. The experimenters hypothesized that in the case when semantically related words, such as "car" and "auto" are presented at the beginning of the list, the subjects would have difficulty recalling the two words in their correct order; when presented closer to the end, which was just before the recall test and the two words were still in the short term memory, the subjects would be able to recall the two words in their serial order more accurately.

In their first experiment Kintsch and Buschke presented subjects with lists of sixteen words. There were two types of lists: one type consisted of eight pairs of synonyms(such as "car" and "auto") presented in a random order, and the other type consisted of sixteen unrelated words. The subjects were instructed to recall and report back the words in the order they were presented on the lists. The results of the experiments supported their hypothesis: The semantically related words caused difficulty in recall only when they were presented early on the list and were not confused when presented late in the list.

In their second experiment Kintsch and Buschke exchanged the synonym pairs for pairs of words that were acoustically related. The results showed that acoustically related words interfere with each other when presented in the end of the list, but not when presented early in the list. These experiments showed that semantically similar words did interfere with each other in the long-term memory but did not interfere in the short-term memory, and acoustically similar words interfered in the short-term memory but not in the long-term memory. It allowed the experimenters to conclude that the dominant mode of the representation in the long term memory is semantic. However, another type of experimental procedure provided direct evidence of the acoustic store of information in LTM.

Nelson and Rothbart (1972) conducted an experiment on elicitation of acoustic information in the long term memory. The basic procedure of the experiment was to present information to a subject and after the subject memorized the information, let some time pass. The purpose was to determine whether the remaining information included acoustic patterns or not. If the answer was positive, then this result could give evidence for existence of the acoustic store in LTM. Nelson and Rothbart presented subjects with a paired-associate task. Subjects learned pairs of numbers and words, such as “37-doe” or “29-cat”. The subjects were required to learn the pairs carefully enough to be able to produce the response (a word) when given the stimulus (number). When the subjects learned the pairs, they were dismissed to their normal activities for a period of four weeks.

When the subjects returned to continue the tests, they were presented a stimulus (numbers) and the subjects tried to recall the matching response (word). The results showed that many of the responses were forgotten. During the second stage of the experiment, subjects were introduced to associate pairs which were reorganized. For example, if "doe" was not remembered with the stimulus "37", then "37" was re-paired with a new response. The new response could be one of the three types: Identical to the old response, acoustically similar to the old response (37-dough), or unrelated to the old response (37-peach). The subjects relearned the new list of words. The second part of the experiment was to answer the question which of the three types of pairs would be recalled better. The results showed that of the three types of pairing, those that were identical to the original form were recalled best. Thus, responses that were *re-learned* were most accurate. Besides that, the acoustically similar responses were recalled more often than those which were acoustically unrelated. The result of this experiment let the experimenters conclude that acoustic information is stored in the long-term-memory.

2.5.2. How information is transferred to LTM

The most widely proven way of transferring information from STM into LTM is rehearsal (Hellyer 1962; Hebb 1961; Rundus and Atkinson 1970; Rundus 1971). The results agreed that the more a word is rehearsed, the greater the probability of recalling the word. Rundus and Atkinson (1970) conducted an experiment in which subjects had to rehearse lists of words aloud. Subjects were allowed to rehearse the responses as long as it was considered necessary, and the time and number of rehearsals were recorded on tape. From the tape recordings, Rundus computed the average number of times each word in the list was rehearsed and from the free-recall he computed the probability with which each word was correctly recalled. The results clearly showed the difference between the words less rehearsed and the words more persistently rehearsed. The results of recall were higher on the latter. The rate of recognition of the rehearsed words was tested three weeks later. The results reported by the subjects showed that the rate of recognized words was higher on the words that were longer rehearsed.

2.5.2.a. Elaborative rehearsal

Craik and Lockhart (1972) distinguished between maintenance and elaborative rehearsal. Maintenance rehearsal is a mechanical repetition of a stimulus. It creates a transient acoustic code for the information. This code may be maintained in memory for a long time, but it does not necessarily transfer the information to the long term memory. The other type of rehearsal is called elaborative rehearsal and involves creating elaborative code for transferring information from STM to LTM. These codes help to properly organize information and to find the appropriate place for it in the LTM. The codes also help to make the information accessible and retrievable. Craik and Watkins (1973) conducted an experiment which clearly demonstrated that simple rehearsal of information does not automatically transfer it into LTM. During this experiment, a long list of words was read to a group of subjects. From the whole list they were assigned to remember and report back only the last word starting with some particular letter ("p"). Because the subjects did not know how many words starting with "p" would be on the list, the only option they had was to maintain in their STM any given "p" word until it was replaced by another "p" word. For example, the subject had to maintain the word "pear" for a period of five intervening items until it was replaced with the word "potato". "Potato" was immediately replaced by "pen". "Pen" was maintained through twelve intervening items until the end of the list. At the end of the list, the subject correctly reported the word "pen" back. After several such lists were presented, the subjects were unexpectedly given a free-recall test on all the "p" words on the list. With this test, the functions relate the recall probability of a word to the length of time the word has been maintained in the STM. The results showed that the function did not reflect a considerable difference among the words maintained in the memory for a considerable amount of time, such as "pen", and the words which were maintained for a short period of time, such as "potato" ($p = 0.18$ to $p = 0.15$). This observation supported the view that the priority in communicating information into LTM does not directly depend on the time of rehearsal and that there are some different factors that influence this process.

2.5.2.b. Mnemonics as a technique of organizing information

It was noticed and described in Ancient Greece that when organized in certain ways, information becomes easier for remembering and recall than when it is structured randomly. The most simple technique of organizing information is "method of loci", which suggests that remembering is based on how information is located in relation to other information (for example, on a sheet of paper, in a speech, by the name of a book or the author, by the order number) can significantly facilitate its recognition and recall. This mnemonic technique employs visual LTM capacity.

Organization techniques integrate information into some already existing memory network or create some new logical framework interconnected with previous networks existing in the memory (Loftus and Loftus: 64). The central tendency in the process of organization is to discover or create any possible connection between the new information and previously stored information. When this connection cannot be easily established, familiar organizing frameworks for organization are employed. These frameworks help to draw analogies with the previous knowledge and thus to tie it to previous structures. The organization of material for better remembering is called mnemonics.

One mnemonic technique was tested and described by Bower and Clark (1969). This method is known as the *narrative chaining* technique. In this experiment, subjects were divided in two groups and were given lists of unrelated words for remembering. It was suggested to one group that they make a story using the words on the list (a story group). The other group was told to simply memorize the words for recall (a non-story group). Both of the groups showed similar probability in recalling the words shortly after the memorization task. The difference between the two groups became evident some time later. After twelve lists were learned and tested, the subjects were unexpectedly given the initial word of the first list and were asked to reproduce the rest of the words. In this test the "story-group" reported back an average of 93% of all of the words on the list, while the "non-story group" could recall an average of 13% of the list. The simple narrative chaining technique increased retention by a factor of almost seven.

Bower (1973) also investigated imagery techniques along with narrative chaining. Imagery techniques imply paired-association paradigm in which subjects pair unrelated words such as DOG - CIGAR. In testing, the subjects were given the first word as a stimulus and they were to report the second word as a response. One group of the subjects was instructed to memorize the pairs of words through rehearsal and the other group was assigned to learn pairs by imagining visual scenes or mental pictures in which the two words are interacting in some way. For example, in the pair DOG - CIGAR, the subject could think of a dog bringing a cigar to its master. Those subjects who used imagery analysis of the pairs performed almost twice as well as the subjects who merely rehearsed the pairs.

How effective organization is for transferring information to the long-term memory was also observed in an experiment which employed the learning of sequences of numbers (Bower 1970; Bower and Winzenz 1969). In the experiment, 12-digit number strings were read aloud to the subjects. The string 828271182890 was presented as "eight-two, eight twenty-seven, one thousand one hundred-eighty two, eight ninety".

The results of recall were improved through multiple rehearsal. During the second trial the grouping of the digits was changed while the sequence remained the same: "eight, twenty-eight, two seventy one...". The third trial again changed the grouping of the digits, but left the sequence as it was. Each time, the subjects spent about the same time memorizing the new grouping, even though the sequence of the numbers had not changed. It proved that retention was not facilitated by the fact that the sequence of the numbers remained the same. It showed how strongly the way the numbers were organized influenced recall.

Organization of mnemonics in free recall was observed in an experiment conducted by Bousfield (1953). A list of 60 words which belonged to four conceptual categories: animals, names, professions and vegetables was presented to the subjects. Each category contained 15 words organized randomly. The subjects were assigned to remember the words and report them back in a free recall, without preserving them in the initial order. Bousfield observed that when the subjects recalled the words they remembered, they tended to cluster the words into the four conceptual categories. In addition to that, the subjects who were given the category names in the next retrieval test did much better than those who were not provided with this retrieval key. The two observations allowed the following conclusions: Subjects tend to structure the material while learning it in the way it was already structured in their memories, and the retrieval keys that corresponded to the structuring of the material in memory greatly facilitated retrieval.

2.5.3. LTM: Organization and retrieval

It was observed that it was easier to recall information in the way it was learned. If the task is to retrieve the information in a way different from that in which it was stored, the task becomes harder. For example, most people remember the names of months in chronological order; therefore, when the subjects in an experiment were asked to report them in alphabetical order, most of them could not do it as fast as they did in chronological order. This showed that the retrieval cue plays a significant role in how fast and effective the process of retrieval is.

2.5.4. Forgetting from LTM

It is still not very clear if any information can be completely forgotten from LTM, or if there is just information which due to some reason becomes inaccessible. There are two main theories attempting to explain the phenomena of forgetting from the long term memory. The interference theory, which was experimentally investigated by Underwood (1964) explained that people forgot something because other information they have learned prevented the event from being remembered. This other information interfered with the desired information. One view on interference claimed that something learned before the control material interferes and prevents recall. This interference was called proactive interference. Another view pursued the opposite perspective: The material acquired after the tested information would interfere more with recalling (retroactive interference). Proactive was first supported by Pavlov's experiments (1927) and then the results were approved by the experiments conducted by Briggs (1954). They both agreed that with an increasing retention interval, initially remembered information undergoes spontaneous recovery and therefore is more likely to interfere with later memory structures.

The other theory of forgetting as retrieval failure explained the phenomenon as temporary inaccessibility caused by not having the right retrieval cue. The strength of reinforcement of recall by retrieval cue was demonstrated by an experiment conducted by Tulving and Pearlstone (1966). Subjects in this experiment were given a list consisting of category names (for example, “animal”, “fruit”, “crime”) and instances corresponding to each category name (for example, “horse”, “lemon”, “murder”). The subjects were assigned to memorize only instances, but not the names of the categories themselves. When the subjects were asked to write down as many words as they could remember, they received different sheets of paper to write on. One half of the subjects received sheets that were blank and the other half received sheets with category names. The subjects that received sheets with category names recalled more instances (an average of 20 words with blank sheets versus 30 words with category names). Later on, the group of subjects that first had the blank sheets was provided with category names. In this case, the subjects could recall about 28 words from the list. Since the subjects were not shown the original list of words before the second part of the experiment, then it could be concluded that these words were retrieved due to the appropriate retrieval cue. The failure to recall the stimulus words could be explained not by forgetting, but by not having the right retrieval cue. This experiment showed a case in which supposedly lost information has been retrieved. It showed how a correct retrieval cue could make recall more effective.

At present, the theory of retrieval failure is more popular than the interference theory described above. However, the differences in the results in different experiments were not significant and sometimes appeared to be caused by chance factors. That is why both theories still exist today as explanations of the complex processes of forgetting from LTM.

2.6. Two types of retrieving information from LTM: Recognition and recall

Recognition is a type of retrieval of information from memory. In terms of response probability, it was noticed that recognition tests on some types of information stored in memory lead to better performance than recall tests (Postman 1950). It was explained from a common sense point of view: Recalling a subject needs more or less "complete" information about the control pattern. To recognize some stimulus, a subject only needs the information necessary to discriminate the target stimuli from the distracter stimuli (Loftus and Loftus: 87).

The difference between recall and recognition is explained in the way in which information is searched for in the memory. The act of recall was considered to contain two sequential steps: In the first step, the location of the incoming information is searched in memory. The subject searches in the long term memory for the words matching those words on the list. The second step in recall is the recognition test of the located candidate in order to ascertain whether the located word is correct. If the candidate is recognized to be correct, then the search stops and the word is given as a response. If not, then the candidate is rejected and the search continues (Kintsch 1968; 1970 a, b; Anderson and Bower 1972).

In the recognition test, the first step of searching and choosing a candidate word is omitted because the correct information is already available to the subject. This correct information is the best retrieval cue, as far as it is the complete information pattern itself. Recognition is therefore a very straightforward way to test whether the information was stored in memory.

The two functions of memory were investigated by an experiment which used the "tip of the tongue" phenomenon. A piece of popular music was presented to the subject. The subject was asked to name the writer. Though the subject was familiar with the music and with its composer, he could not respond with the name. The subject could report an adjacent part of the information, such as the titles of other compositions, written by the same writer, but could not recall the name of the writer himself. However, when presented with a list of fifteen musicians that included the name of the composer, then the recognition became much easier and almost without an error. In this experiment it was clearly shown that when recall was impossible, the information could be *recognized* immediately after a direct retrieval cue was provided. This experiment confirmed that recognition is an easier form of retrieval because a more complete and straightforward information was used as a retrieval cue (Postman, Adams and Phillips 1955).

One more piece of evidence of the distinction between recognition and recall was provided by an experiment conducted by Estes and Dapolito, 1967. They used an *incidental* versus *intentional* learning paradigm. Incidental learning occurs when a subject is presented with a list of words, but is not told that he/she will be required to report them back. Instead, the subject is simply asked to perform some innocuous task such as rating each word for its frequency. Intentional learning occurs when a subject is informed that a present list of words will be required to be reported back. The results of the experiment conducted by Estes and Dapolito showed that recall performance is strongly affected by incidental-intentional manipulation. The recall performance is significantly impaired in the incidental condition than in the intentional condition.

The affect of learning type on recall and recognition was confirmed in the experiment by Woodward, Bjork and Jongeward (1973). In this experiment a series of trials was presented to a subject. Each trial involved (1) presentation of a word for 1 second, (2) a rehearsal interval that varied from 0 to 12 seconds and (3) a recall test of the word. During the recall tests subjects showed perfect results. After the series of trials a final tests was administered. Figure 3 shows the results of this test, placing memory performance as a function of the length of rehearsal interval.

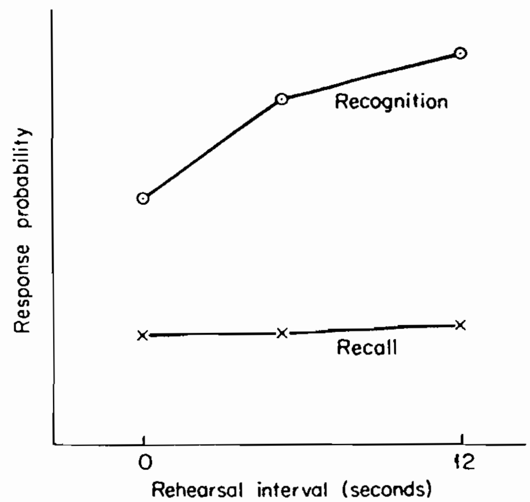

Figure 3. Results of the Woodward et al. (1973) experiment. (Adapted by Loftus and Loftus. 1976)

The results of this experiment support the hypothesis that rehearsal interval does not affect the recall test results. Words that had been maintained in the rehearsal for 12 seconds were recalled no better than words that had only been in the rehearsal buffer for 1 second. In contrast, when the final test was a recognition test, then performance increased as a function of the rehearsal interval. The results showed a difference during testing recall and recognition. When recall test was given, the length of rehearsal interval had no effect on performance. This observation permitted to conclude that merely maintaining information in the short-term memory through rehearsal does not necessarily boost eventual recall performance on the word. Second, maintaining information in the short-term store results in transferring information to long-term memory that is somehow useful for recognition. A tentative conclusion made by the experimenters is that "elaboration" of information in short-term store creates retrieval paths that make recall, more effective than recognition. In contrast, mere maintenance of information in short-term store results in enhancing the retrieval paths that are useful for recognition but not for recall.

Tulving and Thompson (1973: 359) proposed the *encoding specification* principle, asserting that "only that can be retrieved, that has been stored, and that how it can be retrieved depends on how it was stored". Specific encoding operations determine what retrieval cues are effective in providing access to what was stored. So, when recognition was tested, continuous rehearsal was enough for information encoding. In this case, the retrieval cue must be presented as a complete and correct duplicate of the tested information pattern. For a recall test, the information undergoes intentional processing. This means that it is elaborately chosen and consciously organized for providing an effective cue for retrieval. The retrieval cue in this situation will be different from that used for the recognition test, because the same type of cue is not equally effective for both methods of information storing.

2.7. Summary

This chapter deals with behavioral aspects of processing and storing auditory information in memory. The process of learning was described in terms of the Distributed Model of Human Learning and Memory explained by the generalized delta rule of learning. According to this rule, learning is the process of changing the weights of interconnections of units in the human brain in order to adjust it to the external input. The ultimate goal of learning is to have the actual output of the model match the desired output. The changes in the weights of units, activated by external input, determine the way in which the network processes inputs. It was noticed that the changes of weights are processed locally in particular parts of the brain but by the networks that consist of the same type of neurons. That is, each input is processed by only a particular area of the network to which this input corresponds through activating numerous connections with other units and modules.

The second part of the chapter provided a brief review of the processes which auditory information undergoes before it is stored in long-term memory and how these processes are reflected in human behavior. The theories of cognitive processes which provide the flow of information in mind were supported by the description of experiments upon which these theories rested.

Memory was assessed as a unitary multicapacity working structure which according to the level of information processing it conducts was conditionally divided into several stages: sensory memory, short-term and long-term memory. Each of these stages involve different modalities of processing information: Auditory, visual and semantic, activated according to type of perceived information. This division of memory according to levels and capacity allows us to give a closer look to the different stages of the work of memory in information analysis, storing and retrieval. Examining problems of language processing in general, this chapter particularly focused on how *auditory* linguistic information is processed and stored. It was found that auditory information is stored both in STM and in LTM, although each of them has closer connections to one type or the other: Short term memory tends to store auditory information more effectively, while long term memory is more semantic-based.

Any kind of information can get into the long term store either through extensive rehearsal, or through elaborative cognition-based analysis and organizing. The two ways of storing information were found to have different types of retrieval cues. The information stored by means of intensive rehearsal appeared to be better retrieved through recognition, while the information which was stored through analysis and material organization in memory was reported to be better recalled than recognized. Retrieving therefore was found to be closely bounded to the ways of information was stored. Failure to find the appropriate retrieval cue could result in the inaccessibility of information, sometimes confused with information forgetting.

Chapter III. Models of learning L2 pronunciation.

This chapter deals with learning L2 phonology and sound production. I will propose a model of processing and storing a new L2 phonological system in memory and a model of retrieval and production of the L2 sounds. The proposed models are based upon the theories of processing and storing of auditory information in memory discussed in Chapter II.

The model of learning L2 phonological structure (Figure 4) reflects the process of analyzing and storing of L2 phonology and creating a new network of L2 phonological system in memory. A different model (Figure 5) will be proposed for learning production of stored auditory information in the form of acoustic output. The construction of two different models of learning an L2 pronunciation on levels of perception and production reflects the findings in the field of the psychology of memory reported by Howard and Franklin (1987:123) who experimentally showed that separate modalities in memory take responsibility for storing phonological input and for producing acoustic output. The analogous conclusions can be drawn from Flege's (1992) experiments with experienced and inexperienced learners of English as a second language who showed a strong discrepancy between sound recognition and sound production skills. Experienced learners, unlike inexperienced, demonstrated a native-like ability to distinguish between accented and non-accented production of sounds, but did not demonstrate significant difference in sound production from that demonstrated by inexperienced learners.

Acquisition of sound recognition ability and sound production skills in L2 learning corresponds to the two divisions of LTM: Declarative and procedural. Declarative division is characterized as cognitively-oriented and analytical and is composed of subsystems processing various types of incoming information. The procedural division of LTM is expressed through memory systems that do not allow explicit access to the contents of the knowledge. This division can be expressed only in actual performance and is considered to be tied to one of the information processing subsystems in the declarative memory activated by learning tasks. Procedural memory is characterized as slow, more automatic and adapted to incremental learning. Its capacity is limited by the boundaries of the corresponding modality in declarative memory (Squire: 153). Processing and storing L2 phonological information in memory can be related to declarative division of LTM. Since acquisition of L2 sound production skills is tied to and reflects the level of processing and storing L2 phonological structure in memory, we can assume that sound production is corresponded to procedural division of LTM. Therefore, processing and storing of L2 phonological structure and acquisition of L2 sound production skills belong to the two different modalities and require specific input for enabling learning.

3.1. A model of learning L2 phonological structure

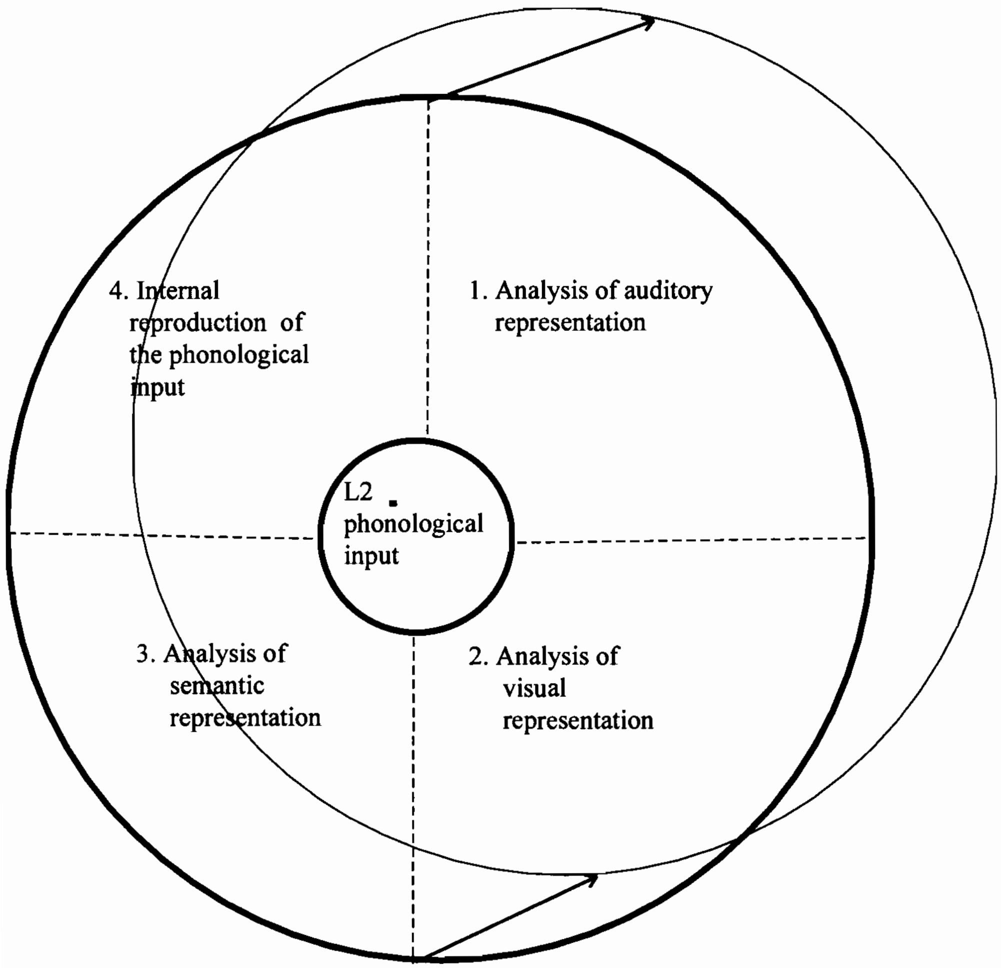

Figure 4. A model of learning L2 phonological structure.

This model reflects the process of creating a new network of interconnections in memory which stores phonology of an L2. The small inner circle reflects the incoming auditory input of an L2. The perceived auditory input through the interconnected units reaches the modalities in memory activated by this input. The modalities are schematically shown as the sectors inside a larger circle separated by dashed lines. The large circle drawn with a heavier line symbolizes STM. The thinner circle which looks like a reflection of the thicker represents the LTM. The two straight arrows moving from STM circle towards LTM circle demonstrate the transferring of information processed by STM onto LTM. Dashed lines which divide STM circles into sectors show the stages of processing of the incoming L2 auditory input by memory modalities and storing of the information in the corresponding STM capacities: phonological (1. Analysis of auditory representation), visual (2. Analysis of visual representation) and semantic (3. Analysis of semantic representation). The order of steps in which the phonological input of L2 is processed agrees with the processing of auditory information proposed by the model of short-term memory auditory store (Chapter II: 2.4.1). According to that model, the auditory information is first perceived and stored in form of raw unprocessed information. The next step is pattern recognition of the maintained auditory information which includes comparing the incoming input to that previously stored phonological information in long-term auditory store. The pattern recognized sounds are then analyzed for their meaning and thus establish the connection with the long-term memory. What is new to the model of short-term auditory store is that in order to be transferred to the long-term memory the sounds must be completely retrievable from short-term memory in their original form. This process is reflected in step four (Internal reproduction of the auditory input).

The process underlying learning L2 phonological structure demonstrated by the model is as follows: Having received activation from the external auditory input, the corresponding module activates other units and modules related to the processing and retaining of the L2 phonological information. The activated by external input units and modules respond with the internal input based on previously stored information. This internal output may be either a representation of phonological system of L1 or of other previously learned languages. When both internal and external inputs localize in a certain area, they start "scanning", or analyzing and recognizing incoming phonological information according to the previously stored categories. Scanning is subdivided into four steps, which come in clockwise order and coincide with STM capacities on the model.

Step one is the analysis of the auditory input of the L2. This step is the starting point in building the structure of an L2 phonological system in the learner's memory. Auditorily perceived sounds of an L2 are analyzed on both phonetic and phonemic levels (Flege: 1992: 571; Odlin: 113). On phonetic level *physical* characteristics of sounds are analyzed including both acoustic characteristics (e.g., the pitch of a sound) and the articulatory characteristics (e.g., how widely the mouth is open in producing a sound). Two languages have sounds that may seem identical but which in fact are acoustically and articulatory different. For example, Chinese velar /x/ is produced with velar friction unlike the English /h/ which is glottal fricative and is pronounced with less friction. Students perceiving the difference in production of these counterparts can modify their production so that their pronunciation comes closer to the standard Mandarin norm. These modifications often do not result in the attainment of target language pronunciation norms, but rather in *approximations* that are neither fully native-like nor target-like (Flege 1992). On phonemic level of sound perception and analyzing learners were noticed to categorize L2 sounds in terms of the phonemic inventory of the L1 (Scholes

1968). For example, Mandarin Chinese diphtong /ɑw/ which is back is categorized by American learners as /aʊ/ which is front, as in "cow". Another example is how Japanese learners of Chinese perceive Chinese nasals /n/ and /ŋ/. Even experienced learners admit that the difference between /n/ and /ŋ/ is extremely difficult for their recognition . At the same time they do not have problems in accurately pronouncing the two sounds.

Although the phonemic inventory of learner's L1 has a strong effect on L2 sound recognition it does not totally impede perception of L2 sounds: The sounds of L2 are recognized by learners as either "similar" or "new" (Flege 1992:573). Similar sounds are those that fall within the range of phonological categories in the L1 or other language learning experiences. New sounds are those previously unlearned and not yet categorized on the phonemic and phonetic levels. A hypothesis proposed by Wode (1981) and Flege (1992) suggested that new sounds in the L2 tend to more strongly attract the attention of a learner than similar sounds, and can be stored in memory more accurately than sounds that are similar to L1 sounds. We can hypothesize that new sounds initiate the development of a new phonological framework for L2 in memory.

Step two in the model reflects the analysis of the visual (orthographic) representation of the auditory input. The orthographic form may be introduced along with the sounds as a standard written form. When the standard form is not received along with the auditory input, learners tend to independently "transcribe" the auditory input using an available written code. The written code can be considered as an orthographic retrieval cue to the received auditory input and serves as a pointer to the L2 phonological information stored in memory. It helps in recognition of auditorily perceived sounds and in distinguishing sound contrasts in L2 by adult learners. In addition to its important role as a universal retrieval cue to the auditory information in memory, the new written representation adds to the weights of interconnections of units in the newly developing L2 phonological structure in the memory of the learners.

Step three in the model is the analysis of the semantic meaning of the auditory information. The purpose of semantic scanning of sounds is to be able to recognize them as meaningful parts of morphemes and words. The ability to analyze the semantic meaning of the L2 sounds shows that sounds are recognized in a context. Speech sounds are related in the mind to the meaning. When recognized in terms of the meaning, the L2 phonological information receives a strong connecting link to the semantic information in LTM and is more effectively stored. Meaning is also a strong retrieval cue in sound recognition and recall. As observed by Martin et al. (1991) subjects show better results in retaining words, rather than non-words. This allowed them to conclude that semantic information in the words helps to keep their phonological representation activated.

The fourth and the last step in the processing cycle is testing the ability to *internally* reconstruct and repeat the sounds of L2. This stage is an important step in learning, as it enables the recall of the whole set of previously stored information in memory. Martin et al. (1994) following Potte and Lombardi (1990) propose that for enabling successful repetition, information must first undergo many levels of processing: from conceptual to phonological. A learner can only realize that he/she has learned something if this knowledge can be recognized or recalled and thus employed in an actual context: while listening to the tapes, participating in classroom activities, or listening to speakers. This stage does not yet imply an ability to produce sounds correctly.

Each of the four succinctly introduced steps of processing L2 phonological information are integrated parts of one process of analyzing and storing information in memory. How closely the steps are integrated in the process can be realized in the case of insufficient input in one of the steps. For example, if the auditory input is not sufficient for starting information processing, then the auditory information gap will be filled through another information input: visual, semantic or merely by retaining in memory a part of information by means of rehearsal. The model demonstrates the multicapacity character of STM, which processes information on different levels: phonological; visual; semantic; and retaining through silent repetition. These levels of processing although being closely integrated, still represent separate functional capacities of short-term memory and were proved to be able to conduct information processing autonomously (Martin, Shelton and Yaffee 1994). Below I will take a closer look at the four steps of processing phonological information in memory.

3.2. Step one: Phonetic and phonemic processing of the L2 auditory input

3.2.1. Phonetic processing

Processing of the L2 starts with the analysis of the auditory information which involves both *phonetic* and *phonemic* levels of analysis (Odlin 1989: 113-115; Flege 1992: 571). The ability to learn phonetically, i.e., to learn to identify the acoustic and articulatory categories of languages and to learn to produce speech sounds closely conforming to the phonetic of the language, is innate in humans. This ability remains intact through one's whole life. However, this ability obtains new qualities by adulthood: First, it is optimized for the encoding and decoding of the sounds in the L1 and second, it becomes resistant to the addition of new phonetic categories, which is partly a result of the optimization process (for discussion see Flege 1990). Flege (1991) claimed that the problems in learning the phonetic system of an L2 (such as the substitution of the sounds of the L2 by the sounds of the L1) by adult learners, occur not as the result of misarticulation or immaturity of the articulation system, because adults achieve full control of the production mechanism. Rather, misproduction may come from *perceptual* misinterpretation of phonetic contrasts in the L2 and from inappropriate categorization of L2 phonetic categories in terms of the categories in L1.

Through a series of experiments with English-speaking adult learners of Spanish, Flege drew the conclusion that adult learners are more likely than children to identify L2 sounds in terms of already established L1 sounds. For example, learners of Spanish, whose native language was English, tended to identify the initial stop /t/ in Spanish words such as *todos* as a realization of the English /t/ category, though their place and manner of articulation are different. Flege (1989) showed that Chinese speakers of English (whose L1 is Mandarin Chinese) can accurately identify word-final tokens of /t/ and /d/ when they contain release bursts, but do so more poorly than native speakers when final release bursts and closure voicing cues are removed (Flege and Wang 1990). Flege suggested that Chinese learners tended to apply the representations used to identify Chinese word-initial stops, which are released, to English word-final stops which are not always released.

Summarizing the results of the research into phonetic learning ability in adults, Flege (1991:251) supported the view in which adults retain the inherent ability to learn the phonetic structure of an L2 without accent, although certain factors may prevent them from optimally using their ability.

3.2.2. Phonemic processing

Flege (Flege and Hillenbrand 1987; Flege and Wang 1990) conducted experiments in which he verified that native and non-native speakers perceive the sounds of the input language differently. Wenk (1985) suggested that the phonology of the L1 causes learners to "filter out" acoustic differences that are not phonemically relevant in the L1. Flege, Munro and Fox (1992) conducted an experiment where they examined how vowel pairs from Spanish or English were perceived by English and Spanish learners. The analysis of the data showed that the native Spanish-speaking listeners use fewer dimensions in judging between vowel dissimilarity than the native English listeners. This finding corresponds to the observation that Spanish speakers use a narrower range of tongue positions to produce Spanish vowels /i/, /a/, /u/, than English speakers use in producing English vowels /i/, /a/, /u/. Spanish listeners used fewer dimensions in identifying English vowels because their native language had a smaller vowel inventory and fewer phonetic contrasts than English. Munro (1990) found that native speakers of a language where phonemic length is a distinctive feature (for example, Arabic), used duration to a greater extent to classify vowels of an unknown foreign language than the speakers of English, who do not have this type of phonemic distinction in their L1.

A set of experiments presented the evidence that L2 learners can develop the ability to detect acoustic properties of the L2, and can establish representations for the L2 sounds in their memory through increased experience. The study conducted by Flege (1988) examined whether foreign accent perception can indeed be based on divergence from L1 phonetic norms, in addition to overt substitutions, and whether non-natives can perceive a foreign accent in the L2 even if they themselves speak it with a foreign accent. Three groups of listeners took part in the experiment: native speakers of English, Chinese subjects who had lived in the USA for about five years, and Chinese subjects who had lived in the United States for about one year on the average. The listeners estimated the degree of foreign accent in English sentences. The English sentences which were offered for evaluation were spoken by a group of native speakers of English and by groups of Chinese speakers, differentiated according to the age when L2 was learned and/or amount of previous English-language experience. The estimation of the degree of foreign accent ranged on the scale of the response box between endpoints marked "strong foreign accent" and "no foreign accent".

Two assumptions have been made to predict the results of this experiment. The first assumption predicted that experienced and inexperienced Chinese listeners would show no difference in judging the foreign accent based on perceiving the substitutions of English sounds by Chinese sounds (sentences, produced by beginning Chinese learners), as it would be equally easy for both groups of Chinese listeners. The other assumption predicted that if foreign accent perception is influenced by segmental divergence from phonetic norms (sentences produced by experienced English talkers), better performance should come from experienced rather than from inexperienced learners, because experienced learners would develop *central representations* (or mean formant frequencies; for discussion see Flege 1992:583-587) for how English categories ought to sound (Flege 1991:258). The results of the experiment clearly supported this latter assumption. This implied that L2 learners develop central representations for English phonetic categories. The ratings accorded to the speakers by various groups of listeners agreed with the expectations. Sentences spoken by native speakers of English and by Chinese subjects who had learned English at an average age of 6-7 years ("child learners") received higher ratings than sentences produced by Chinese speakers who had lived in the USA for about one year. An important difference was noticed in the data between the three listener groups. Native English-speaking listeners differentiated native and non-native speakers to a significantly greater extent than the experienced Chinese listeners, who, in turn, differentiated native from non-native speakers to a significantly greater extent than the inexperienced Chinese listeners. If one were to assume that subjects in both listener groups responded in the same way to segmental substitutions, these results could suggest that the experienced Chinese listeners had a better notion of how the English sentences ought to have sounded than the inexperienced Chinese listeners. They may have done so only by having established phonetic norms for sounds in the English sentences.

The results of this experiment clearly demonstrated a dichotomy between speech production and perception. The subjects in the inexperienced Chinese listener group (Group A) participated as speakers, and subjects in the experienced Chinese listener group (group B) were drawn from the speaker group. Even though the Group A and Group B speakers produced equally strong foreign accents (as assessed by native English speaking listeners), the group B listeners were nevertheless better able than the group A listeners to differentiate native from non-native speakers of English. This shows that the ability to gauge the degree of accent in L2 sentences may increase more rapidly (or to a greater extent) than the ability to produce L2 sentences authentically (Flege 1991:261). Thinking about the representation of speech production and perception in memory in terms of the results of this experiment, we can first find out that the ability to correctly identify sound contrasts depends upon a growing experience in speech production; on the other hand, the development of speech production does not directly depend upon a growing ability in speech perception. It can be concluded, then, that a special input stimulus is necessary for the development of sound production separate from that, necessary for learning sound recognition. The issue of interaction between speech recognition and production will be discussed in more detail in section 3.5 of this chapter. Here we are primarily concerned with how recognition memory is structured and what role it plays in learning L2 sound production.

According to a widely supported hypothesis (Wode 1981; Flege 1991), unless non-natives develop L2 phonetic categories that are the same as native speakers', they may be unable to produce L2 sounds authentically. This hypothesis implements the assumption that in sound learning, recognition of sounds of the target language precedes the development of the articulation programs needed to implement phonemic categories. It implies that at a certain stage of learning, non-natives may develop perceptual norms like native speakers, but may not be able to produce them authentically.

3.2.3. Learning "new" versus "similar" sounds of L2 phonological system

Part of the studies on acquisition of L2 phonological structure focused on the perception and production of "new" vs. "similar" sounds of the L2 by learners. The results of these studies provided preliminary evidence that new sounds may be learned better than similar sounds (Flege 1991; 1992). The distinction between new and similar sounds was discussed by Delattre (1962; 1969) and later by Wode (1978) and Flege (1991, 1992). Wode (1978:114), speaking about the phonemic processing of the sounds of the L2, noted that learners match phonetic elements of the L2 to their L1 "grid". While processing the L2 acoustic information, the input is "scanned". The phonetic categories that fall within the similarity range are judged to be equivalent to categories of the L1, and the categories, that do not fall within the similarity range are judged as "non-equivalent". The categories that were judged as equivalent to the categories in the L1 are substituted by equated elements in the L1, and the categories that were found non-equivalent will undergo further analyzing. Flege (1992:575), conducting experiments with English learners of French, found out that English learners could achieve authentic production of the phoneme /y/, which is new for native English speakers, but could not control the production of the similar sound /u/, which is not fronted in French like its English counterpart. From a series of experiments, Flege hypothesized that L2 learners are unable to establish additional phonetic categories for similar L2 phonemes because they are equated with L1 phonemic categories and, as a result, stop further processing. New categories, in contrast, form new phonemic categories and can be pronounced accurately.

3.3. Step two: Analysis of visual (written) representation of L2 phonological structure

In spite of the diversity of written representations of sound systems among the languages of the world, written language was created as a representation of the spoken language in a visual form. The close relationship of the written and spoken forms can be traced in the influence of spoken language on the written form. In the languages which adopted alphabetical writing, for example, the Slavic group of languages, the historical and geographic development of the spoken norms caused changes in the original common alphabet and in the creation of new letters and markings. As a result, such new alphabets as Russian, Polish, Ukrainian and Bulgarian were formed. The orthography was also gradually adjusted to the norms of spoken languages which resulted in designing new letters and symbols and in eliminating symbols reflecting sounds which were no longer in use. The changes of old scripts conserved the features of gradual historic and geographic developments of spoken norms and preserved information necessary for reconstruction of old pronunciation.

In languages with a morphographic writing system, like Chinese, the phonological system was represented in a different way from alphabetic languages. Nevertheless, the phonological origin of the writing system can be recognized. The phonological origin of writing is reflected in the way in which a large number of Chinese characters have been designed: the principle of borrowing characters for representing other words with a homophonous reading but different meaning. This principle of character design on the basis of the principle of phonological resemblance of characters was called *jiajie* or "loan words. These are graphs, which originally were designed for one word, but later were borrowed for another word because of the similarity in reading with the original word. Another principle of character design employs two (or more) components in one character, one of which gives a clue to the *semantic* category of the word and the other-- phonetic component-- suggests *reading*. For example, the graph for the word *hé* "river" consist of two graphs: one graph means water on the left side and the other graph on the right side is pronounced *kě*. This right-side component suggests reading of this composite graph. According to Norman (1988: 68), phonetic component only occasionally coincides perfectly with the pronunciation of the graph in which it is used.

In contrast to the languages with alphabetic writing, historical changes in the spoken norms do not affect the appearance of these characters. Norman suggested that the parameters of phonetic usage are sufficiently narrow to probe valuable information about the phonological make-up of Han and pre-Han Chinese. Because the sounds are represented indirectly, through using phonological similarity of one morpheme to represent another morpheme with similar or the same pronunciation, significant variations in local and dialectal pronunciation norms has developed within the same written code. The traditional Chinese writing system as a retrieval cue of the phonological representation of the language, therefore, presents a number of limitations. First, Chinese characters do not give a direct cue of how this character must be pronounced: Unless one knows the character, he/she cannot pronounce it and can only guess how it actually is pronounced. Second, one character represents the whole syllable, e.a., a group of sounds and not one sound as it usually is in alphabet writing. This often results in inaccuracy of retrieving its phonological representation by learners, as one representation can easily be confused in memory with another which is similar acoustically or articulatorily. For example, the syllable *chǎo* is frequently confused by learners and mispronounced as *shǎo*, which is acoustically similar. On the other hand, Chinese traditional written code has an advantage over the Romanization system in retrieving phonological information: Chinese characters are the authentic representation of the phonological structure of the Chinese language and are directly related to the sounds of Chinese in memory. This excludes any opportunity of interfering of L1 phonological structure in retrieving process and provides strong retrieval paths to the phonological system.

An alphabetic writing system is an effective tool for learning phonological system of an L2. In order to prevent the quick decay of the incoming L2 auditory information and to maintain its longer retention, adult students tend to "visualize" sounds by transcribing them with an orthographic code. Written representation of the L2 creates a logically related system of retrieval cues which serves as a pointer to the stored in memory L2 phonological information. Flege (1992) who observed that an adult learner's auditory perception and processing is influenced by the categories in sound system of the L1, claimed that difference in perception changes L2 actual phonological structure in learner's memory. Alphabetic writing system through a letter-to-a-sound representation the phonological structure of L2 helps adult learners to realize phonetic and phonemic differences of L2. Learners obtain a visual aid for distinguishing the L2 sounds by matching the auditory and visual data. I will present an example from my experience to explain this statement: A beginning learner of Mandarin, whose L1 was Cantonese, did not show that he clearly perceived the distinction between /ʂ/ and /s/, and kept substituting the sound /ʂ/ by the sound /s/. Rehearsal and correction of articulation did not give long-lasting effect: Several minutes later the student, repeating after the instructor, returned to pronouncing the wrong sound. Then the student was asked to spell the word *lǎoshī* -- "teacher" as he heard it in *pinyin*. The student wrote: *l ǎ o s ī*. After the "spelling" was corrected and the correct pronunciation rehearsed, the student showed noticeable improvement in attending to this sound distinction. Thus, orthographic written code played an important role in helping the student make a distinction between /ʂ/ and /s/.

In my practice of teaching English to Russian students, one situation was very common: Beginning learners, who were not yet very familiar with the writing system of the L2, used their L1 writing code for writing down the new words used in the class. When asked to reproduce the new words, the students used their notes and pronounced the L2 words with the strong interference of the phonological structure of the L1. This example from the classroom, where teaching of the L2 relied upon the perceptual memory without employing the written representation of words, demonstrated two important issues for introducing written code in sound learning: First, that adult learners tend to record the words in written form to facilitate later retrieval. Second, when using the written representation of the L1 for L2 sounds, learners automatically substitute the L2 sounds by the L1 sounds. This happens because the L1 sounds are connected with the L1 writing code in memory. This observation shows that the sounds of the L2 cannot be retrieved accurately by using the writing code of the L1. This allows us to assume that L2 and L1 sounds have different retrieval paths, and consequently, have separate phonological structures in memory.

Another example will strengthen this assumption. Very often I receive e-mail, where the text in Russian is transcribed with the letters of the Latin alphabet. Reading the messages, I catch myself pronouncing Russian words with a strong English accent, being influenced by the written representation. This considerably slows down my processing the information. I have to force myself to pronounce the English letters with a Russian accent in order to understand the text. These experiences permit us to conclude that the written representation of the L2 is a part of the L2 phonological structure in memory because it has retrieval paths different from that of L1 written code. Therefore, for enabling an accurate retrieval the written code must be associated in memory with an L2 phonological structure, e. g., be an authentic, related written code for L2. Introducing the authentic written code for representing L2 sounds is an important step in forming the autonomous system of the L2 phonological structure.

The written form of L2 sound representation, however, must not substitute or precede the auditory input when new sounds of L2 are presented. Sounds must first be retained in the echoic memory in their authentic acoustic form in order to enable analysis of the written representation of the sounds (Lukatela and Turvey 1994). The sounds of L2 stored in the STM are then analyzed according to their written representation. Written representation as a retrieval cue must already have the auditory information previously stored in memory in order to be able to retrieve it, otherwise retrieval would not occur. Human memory being more creative than any artificial intellect, would not respond: "The retrieval path is invalid" and stop there, but would suggest an alternative interpretation. The phenomenon of the learners who were never (or rarely) exposed to spoken English and still learn how to read and write in English would seem to contradict the hypothesis supported by the experiment conducted by Lukatela and Turvey (1994), in which they showed that visual lexical access is initially phonological. However, the self-taught learners of English with zero (or near zero) spoken ability do pronounce words when they read or write them: they reproduce the words letter-by-letter, using the names of the letters as they were introduced in the English alphabet or the IPA, having no other stored auditory information to retrieve. For example, beginning learners of Chinese who do not receive a sufficient auditory input, sometimes instead of */ts/* in a Chinese word *cai* will pronounce */k/* using a rule that in written English the letter "*c*" in a word initial position is read as */k/*. Thus, we can conclude that the writing system retrieves the information from memory, through recoding it first into the auditory modality. This observation supports the view that when introducing the phonological information of the

L2, the auditory input must precede the visual input, but the visual input in writing must follow the auditory input for enabling a more accurate processing and storing. The auditory and visual modalities are closely integrated in analyzing the auditory input of the L2 in the case of adult learners. Even if the auditory input must be considered as the prime condition of learning a sound system and the written representation as the auxiliary one, a related written representation is an extremely important factor for complete sound processing in memory.

3.4. Step three: Analysis of semantic representation of L2 phonological structure

Semantic analysis represents another step in the L2 sound system processing. The subjects of this analysis are the units which bear semantic meaning --words, compounds, sentences, etc. Semantic analysis is involved in acquiring the phonological structure by using the words' meaning. It was observed by Martin et al. (1994) that when auditory information is analyzed from the semantic point of view, it helps to retain the auditory information longer in memory. This means that when sounds are pattern recognized for meaning they obtain a connection with a semantic capacity in the long-term memory. The sounds have a stronger opportunity to be transferred to the long-term memory. Observations show that when learners know a word's meaning it gives them a better opportunity to concentrate on how the word is actually pronounced. Flege (1991:260) found that experienced Chinese learners of English who had lived in the United States for an average of five years, unlike inexperienced Chinese learners, who had lived about a year in the country, could distinguish very well the difference between the speech patterns produced by native and non-native speakers of English. They could also distinguish between the patterns with a strong foreign accent from patterns with a light foreign accent. This observation allowed Flege to conclude that experienced learners develop "central representations" of English segments within which the unaccented production is represented. How could this range of central representations be developed? The experienced group of learners had an average of five years of experience living in an English language environment, they had a longer period of actual functioning in the target language which implies extensive meaningful interaction with native users of the target language. Namely, the knowledge of meaning of words helps learners in communicating with different regional and dialectal groups of native speakers of L2 and allows them to collect and store data of possible pronunciation modifications for each sound. This cumulative data of a diversity of pronunciation (different social groups, or regional varieties) formed a *central phonological core* for each phoneme in the memory of learners. This phonological core, or "*central representation*" (Flege 1992:589), allowed them to distinguish between accented and authentic spoken norms. Thus, we can assume that learning L2 phonology can be conducted through its application in a semantic context, through learning words and speech patterns. This process is known as a natural approach with top-to-bottom processing (Hammerly 1991:31).

However, when learning is focused on communication of meaning, accuracy in pronunciation may be overlooked and then acquisition of L2 phonological structure is impaired. An example, illustrating this statement can be found in Odlin (1989: 114): The uvular /ʀ/ of Parisian French and the retroflex /ɽ/ of American English have very different phonetic properties, but learners have problems to recognize the phonetic difference of the sounds in such word as *route* and confuse the two sounds because the two words in English and French have similar semantic cues. (Odlin: 114). Learning to communicate the meaning before the control over the accuracy has been achieved, therefore, can present an obstacle in acquiring accuracy of recognition and pronunciation.

3.5. Step four: Internal phonological production of L2 auditory input as a part of learning L2 phonological structure

The fourth step in the proposed model of learning phonology of L2 is achieving the ability to covertly produce the L2 sounds by learners. Covert speech, or, egocentric speech, as identified by Vygotsky (1962), is a way of learning language by silently pronouncing words . This way of learning words was observed when children learn their first language. Usually this way of learning disappears at the age of elementary school. By this age children are usually able to learn and produce new words without internally pronouncing them, as well as to think without pronouncing words, e.g., they develop the outer and inner speech in their language, which is a feature of a higher level of language acquisition (Vygotsky: 126). The stage of covertly pronouncing words in the language is therefore an important stage in developing memory for the L2. Using the internal pronunciation of words is involved in retaining and storing auditory input and also for recall and reinforcement of the auditory material retained in memory. During silently pronouncing a word, the learner does not hear how he/she actually produces the word or pattern, but reconstructs the last production of a word and activates the entire set of information about this word that is retained in his/her memory. The internal production is focused particularly on activating the previously stored auditory and other types of information in the memory and, unlike the overt production, does not affect the newly stored L2 auditory input.

Experiments on the role of voicing and silent rehearsal in word retention were described by Stevick (1976: 31). In this experiment one group of subjects was required to pronounce aloud the words in a list, while another group read the same items silently. The recall by both groups of subjects was compared. The group which used silent way of rehearsal showed better results, than the other group. It was concluded in the experimenter's report, that the requirement of active vocalization could interfere with effective coding operations. Active vocalization could also demand special attention which would lessen the attention capabilities that are necessary for encoding. The advantages of covert production were found to be in longer retention intervals, when the information in the echoic memory has dissipated (Tell and Fergusson: 349). These observations enabled experimenters to conclude that the non-voiced items were processed to a deeper level than the voiced ones.

The two experiments mentioned above were exploring retention in the working memory of the group of items in an L2. The same assumptions hold true for retention of the words in an L2 (Stevick: 34). The silent retention of auditory input activates the capacities corresponding to the auditory modality (phonological, semantic, visual). This observation may be explained by the fact that when produced aloud, words create additional auditory input which activates echoic memory and thus interfere with the previously stored authentic auditory information. Therefore, for better reinforcement of the phonological memory and for more accurate retention of authentic auditory input silent production is more effective than vocalizing.

3.6. Model of learning L2 sound production and phonological memory

It was noticed that in spite of the fact that adults gain better control of the articulation system than children do, in most cases they cannot achieve the same degree of identity with native sound production that children can. Flege (1992) proposed that learning deficit in adults was not caused by their inability to change stabilized articulation mechanism, but by the perception at which adult learners tend to categorize the L2 sounds by means of phonetic categories in L1. For example, American learners of Chinese usually categorize retroflex /ʂ/ of Mandarin Chinese as a postalveolar fricative /ʃ/ which is introduced in their L1. This phenomena prevents adult learners from establishing new phonetic structure for L2 in memory, which means that learners also will be unable to develop new articulation mechanisms and produce sounds of L2 authentically. The inability to perceive the sounds of L2 as they are, without categorizing them in terms of previously stored phonological data of L1, is explained by the fact that the phonetic space in memory of adults was "filled" by phonetic categories of L1 at a greater extent than children's and cannot be filled more (Flege 1992: 595). The phonetic space in memory is getting filled by outward expansion of category boundaries, which occurs while children growing older learn to correctly identify central pronunciations and also "ambiguous stimuli", e.g., sounds, produced by an ever wider variety of speakers differing in vocal tract size, dialect and characteristic speaking rate. The hypothesis of filling up the phonetic memory space by adult age was proposed by Flege on the basis of experimental material by Krause (1982 a,b), who investigated children's production and perception of vowel duration on the contrast between voiced and voiceless stops in word-final position. He drew a conclusion that both perception and production demonstrated stabilization with increasing age (Krause 1982a: 25). Another investigation by Scovel (1988) discovered that when children grow older they can better gauge the degree of foreign accent in realization of a particular phonetic category of their L1, which may confirm that they stabilize their perception categories around central phonetic representations of L1.

The opinion that L2 phonological structure cannot be acquired in adult age because phonetic space was filled up by L1 categories raises a question what type of learners of L2 participated in the experiments described by Flege. According to the description, the participants were of two groups: experienced learners (those who lived in the USA for five years) and inexperienced learners (those who lived in the USA for about one year). However, it was not specified what type of instruction the learners experienced, if it was a formal ESL type of instruction where pronunciation is learned systematically or if it was a naturalistic learning in actual communication in the target language environment.

Experience shows that, there is a crucial difference between the these two categories of learners, and the difference is especially noticeable in their acquisition of the phonetic categories of L2. According to the Distributed Model of Human Learning and Memory, guided or supervised learners can develop authentic or authentic-like pronunciation, while unguided learners may never acquire accurate pronunciation in L2. The importance of strong feedback in learning pronunciation is evident from Flege's hypothesis about the inadequacy of the perception of L2 phonology by L2 learners. Inadequate perception by learners leads to an uncontrolled substitution of L2 sounds with L1 sounds or their modifications. If one were to rely on the Flege's hypothesis, then beginning learners cannot merely correct their pronunciation by themselves, because they perceive the L2 sound categories in terms of L1 sound categories and, therefore, produce the perceived sounds in terms of phonetic categories in L1. Thus, beginning learners, who have not yet identified the L2 sound categories, need constant feedback which would demonstrate to them the different phonetic categories of L2 and gradually change their perception. The change in perception towards native-like would mean that a new structure was established in phonological memory. The creation of the new phonological structure is an important step for changing sound production towards native-like, as it provides a degree of control over the produced sounds.

Change in perception does not imply that articulation will also immediately become authentic-like. Perception and production of sounds differ functionally and belong to different modalities in LTM: declarative for sound recognition and procedural for sound production. (Squire 1991:152) The change in the articulation would require additional input, that is qualitatively different from that necessary for developing accurate perception. Below is demonstrated the model of sound production which reflects the interdependence and interaction between the two memory modalities in the process of learning.

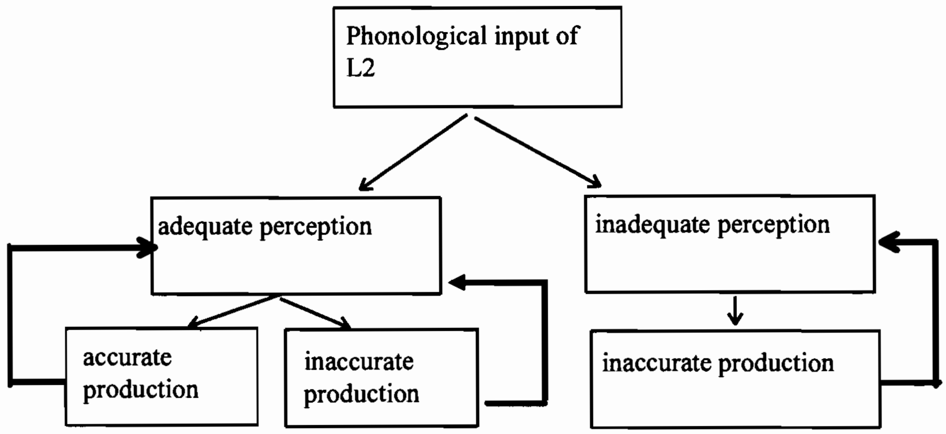

Figure 5. A model of learning L2 sound production.

The articulation system of a learner receives phonological input from L2 which originates in memory and responds to it with output. The output, especially at the beginning stages of learning, will receive a strong influence from the L1 patterns of articulation and would interfere with the phonological data of L2 stored in memory. Multiple production of the L2 patterns without authentic feedback and correction will modify the phonological categories in memory to the L1 phonological categories and will result in losing their original authenticity. Thus, the process of learning sound production is crosscurrent: Phonological structure in memory stimulates changes in the articulation system for pronouncing the L2 sounds, and at the same time articulation mechanism which from force of motor habit equates production of the sounds of L2 with the sounds in L1 imposes the L2 phonological memory and adjusts L2 sounds to similar L1 sounds. This process of mutual adjustment between articulation mechanism and sounds in memory may reach a balance and develop intermediate pronunciation norms which will be equally distant from both languages. Therefore, persistent auditory input of authentic pronunciation along with consistent correctional feedback is crucial in breaking this balance. Persistent phonological input reinforced with sufficient theoretical explanations of the nature of the L2 sounds and effective error analysis will result in the gradual adjustment of the articulatory memory for the phonetic categories of L2.

There is another type of input which is psychological by character. Stevick (1976: 59) proposed that inability to produce sounds of L2 authentically is not physiological, rather it is mostly psychological. He considered that a proper frame of mind which will help to imitate the L2 native speakers' pronunciation is the most important issue in learning pronunciation. Learning pronunciation by mere imitation is an oversimplified view of the process, but the psychological input is very important in learning pronunciation. Vygotsky (1962: 130) asserted that internal speech is personal, while outer speech is a social phenomenon. The desire to learn authentic pronunciation by learners at a great extent depends on the generally accepted attitude to the target culture and language in society in a broad sense and particularly in the classroom. The general attitude of the social group and the popularity of the target language influences how close to the target language a learner would like to sound. This means that the learning of pronunciation is strongly influenced by the mental state of a learner, which is determined by the degree of motivation and integration into the target culture.

3.7. How the Distributed Model of Human Learning and Memory supports the proposed models of learning L2

The models described in this chapter can be actually verified only by practical application in the teaching of L2 sound system. The recommendations on the application this model in the classroom will be presented in the next chapter.

The main thesis of the proposed model of learning is that both similar and new L2 sounds can be learned authentically by adult learners in the case of supervised instruction guided by proposed models. This thesis will be verified through the general delta rule of learning in the Distributed Model of Human Learning and Memory by McClelland and Rumelhart (1986) as described in Chapter II.

According to the Distributed Model of Human Learning and Memory, learning is a process of developing weights in the interconnections of memory network of interconnections that enables it to respond appropriately to a set of input stimuli. Learning is made possible owing to the ability of interconnections to adjust their weights to match external input. This phenomena was called delta rule of learning (McClelland and Rumelhart 1986: 180) and was presented in the formula:

$$\Delta w_{ij}=\eta\delta_i a_j \qquad (1)$$

where Wij is the change in the weight to i from j, η is a global strength parameter which regulates the overall magnitude of the adjustments of the weights, and δ is the magnitude which determines the size and direction of the overall change that needs to be made in the internal input to the unit. The delta rule of learning utilizes the discrepancy between *desired* output pattern and *actual* output to improve its weight during training phase. The activated network receives an input pattern, on the basis of existing weights generates an actual output, compares it to the desired output, and then changes each weight based on the *difference* (or *error*) at each output unit. This process is also called an error correction procedure and is a prototypical example of a supervised learning.

The error for the entire output pattern is referred to as pattern sum of squares, *pss* (Bechtel and Abrahamsen 1991:74). The difference between the desired output of a unit (*u*) (d_u) and the actual output of unit (a_u) is computed for each unit. The difference is squared and the overall squared differences of output units are added together:

$$pss = \sum_{u} (du - au)^{2} \qquad (2)$$

A system can obtain the total sum of squares (*tss*) by simple summing of the *pss* values. The sum indicates how much room for improvement remains and evaluates the progress of the entire set of input-output cases.

The process of learning according to the delta rule requires that each weight in the network be changed according to the following equation:

$$\Delta weight_{ui} = lrate \ (d_u - a_u)a_i \qquad (3)$$

Each weight is regarded to be adjusted individually in the delta rule in order to reduce the total error. When the difference, or error, is equal to zero, then none of the weights feeding into it is changed and the desired output has been produced. When there is an error and the input was not zero, the equation causes a change in the weight in the direction that would reduce the error. The error correction procedure of the delta rule of learning shows that progress is possible through gradual adjustment of the actual output to the desired output. This model also shows the advantage of supervised learning in comparison with unsupervised learning (Bechtel and Abrahamsen 1991:86): In supervised (reinforced) learning an activated network is explicitly told what output was desired for a particular input and the actual output is compared to the desired output. In unsupervised learning the network classifies a set of inputs without feedback. Therefore, in unsupervised learning the change of weights would not necessarily go in the direction of desired output. It can be concluded that learning is able to achieve desired output through gradual adjustment weights of actual output to the desired output. This adjustment is due to error correction in supervised learning. The delta rule of learning, therefore, supports my hypothesis that L2 sound system can be learned accurately. Both familiar and new sounds of L2 can find its reflection in memory if undergo thorough and consistent error analysis and correction. Thoroughness and consistence of error correction is reflected in the ability of an instructor to realize all sources which may cause misproduction of L2 sounds (using the terminology of Distributed Model of Human Learning and Memory, to decompose total error (*tss*) into smaller units (*pss*), and to correct each error individually. The extensive authentic input is necessary to enable a learner to compare actual production with desired production and make adjustments.

3.8. Summary

In the model of learning L2 phonology proposed in this chapter, I described four major individual components, which while being closely integrated in learning, nevertheless follow each other in succession order to form a new structure of L2 phonological representation in memory. These four steps are as follows: L2 auditory input analysis and storing; analysis and storing of L2 visual representation of sounds; semantic analysis of L2 sounds and words; and internal rehearsal and retention of L2 sounds. The model reflects the multicapacity structure of memory which is involved in learning L2 phonology and which is gradually activated by L2 input. To verify the importance of the succession order of the four steps of information processing, the reverse order of stages can be assumed. The learner starts learning L2 with perceiving and storing meaningful units -- words and phrases in order to be able to start to communicate as soon as possible. A learner's attention is for the greater part focused on remembering variety of meanings of words and phrases, grammar patterns and suprasegmental features. Words that are learned extensively are stored in phonological approximation to its original form and the L1 phonological system is used to substitute sounds of L2. In this situation the hypothesis supported by Flege and Wode is completely appropriate: The learners will start using L1 sounds for familiar sounds in L2 and will develop new categories for new sounds of L2. This is due to insufficient feedback which provides comprehensive analysis of phonological errors. The method of involving students in L2 communicative interactions by using as much vocabulary as possible assumed that students can become linguistically proficient by communicating using many errors at first and gradually overcoming their linguistic inaccuracy (Hammerly: 124). However, the use of words and phrases in meaningful communication before the threshold of phonetic and phonological control has been reached results in far more errors than can be effectively corrected, and this may result in permanent adoption of faulty habits (Hammerly: 124). Hammerly proposed that mastering L2 involves linguistic and communicative integration every step of the way. His view on a balanced approach in achieving communicative and linguistic competence in general supports my hypothesis that teaching to pronounce L2 sounds authentically has to involve persistent phonological control with comprehensive error analysis and immediate correction. However, the "balanced approach" term has a very broad sense and it does not set priorities which could be employed in teaching practice. Investigation of memory structure reveals that when one capacity of memory is activated, others are impaired. It means, that when meaningful communication is practiced, sensitivity to phonological representation in memory is impaired, or less attended and it results in misproduction and unconscious substitution of L2 sounds by L1 sounds or their modifications. The model of learning which was proposed in this chapter focuses on the developing phonological store as a priority in L2 phonology acquisition while visual and semantic representation of sounds have an auxiliary role in developing accuracy in L2 phonology.

Chapter IV. Teaching Mandarin Chinese pronunciation: Review and proposals

This chapter deals with the application of the cumulative knowledge of memory and learning to teaching Mandarin Chinese pronunciation. In the previous chapter we discussed the problems of mispronunciation of sounds by learners and tried to reveal the reasons for the development of a foreign accent. We proposed and described models of learning pronunciation of the L2. Analyzing the nature of the pronunciation errors made by adult learners as observed and described by Flege (1991, 1992), Wode (1981) and Hammerly (1991), we came to the conclusion that native-like pronunciation can be acquired by adult learners, if the necessary input and feedback are provided. The reason why many adult learners do not develop accurate pronunciation skills was revealed to be not in the physiological inability of adults to develop new phonetic categories, but in their receiving insufficient input and/or correctional feedback. This observation is reflected by the delta rule of learning depicted in the Distributed Model of Human Learning and Memory (McClelland and Rumelhart 1986; Bechtel and Abrahamsen 1991). According to Flege, (1992:570) phonetic errors are originated when sounds are perceived and processed in memory, but not during the stage of sound production. It was further experimentally confirmed that the ability to correctly recognize sounds of the L2 does not imply the ability to produce sounds accurately. Sound perception and sound production proved to be related to separate modalities in memory, each requiring individual input and feedback for its development.

In the present chapter we will propose guiding principles for teaching Mandarin Chinese pronunciation to adult learners. These guiding principles were designed upon the analysis of the present approach of teaching pronunciation in The Ohio State University in the Department of East Asian Languages and Literatures in the classroom and in the individualized instruction modes and from the observation of the most common errors made by learners which provided the necessary platform for analyzing the reasons for mispronunciation.

This chapter consists of three major parts. In the first part the most common errors made by learners of Chinese will be described and analyzed. In the second part we will discuss the possible preconditions which cause the acquisition of wrong pronunciation. In the third part of the chapter, guiding principles for teaching Chinese pronunciation will be proposed.

4.1. An analysis of the most common mispronunciations made by learners of Mandarin Chinese.

Common mispronunciation errors can be noticed among different groups of learners of Chinese as a second language. To conduct the close analysis of the mispronunciation patterns is very important in developing the strategy of dealing with this problem. For instance, a series of pronunciation drills and exercises can be developed for each particular group of learners whose L1 interference causes special type of pronunciation errors. The analysis of mispronunciation will enable to explain the reasons for mispronunciation more consistently.

Here we will provide some examples of mispronunciation patterns made by foreign learners of Mandarin Chinese. The examples enlisted below is the result of the personal observations made by instructors who are native and non-native Chinese speakers. The instructors have an extensive experience teaching Chinese in the Department of East Asian Languages and Literatures at The Ohio State University. The description of how the segments and tones are pronounced according to the norms of Mandarin Chinese is based on *A Practical course in Chinese phonetics* (1982), Paul Kratochvil's *The Chinese Language Today* (1968) and Raymond Huang's *Mandarin Pronunciation Explained with Diagrams*.

The pronunciation errors were reported to be of two categories: misproduction of tones and misproduction of segments. The Tables 1 and 2 will review most common mispronunciation patterns and analyze the reasons for mispronunciation.

Table 1. Segmental errors made by foreign learners of Chinese. (Initials)

Target sound IPA	pinyin	Most common misproduction	Reasons for misproduction
ts	(z)	s, tʃ	This unaspirated voiceless affricate is pronounced in two steps: First, placing firmly the blade of the tongue to the upper teeth and the side rims of the tongue to the sides of the upper teeth, thus compressing air inside the mouth. Second, release the closure by slowly lowering the tip and the front of the tongue and letting the air flow out with a sibilant noise. Students mispronounce this sound by not compressing the air in the mouth and letting it flow through the passage gradually, thus producing the sibilant noise without the release burst, which resembles the sound /s/. This mispronunciation was noticed predominantly among American learners of Chinese. Pronouncing this sound noticeably voiced, while the sound is considered as voiceless is another misproduction pattern of this segment. This sound is also misproduced by replacing /ts/ with another affricate /tʃ/. The reason of misproduction is in the wrong position of the tongue. The sound is misarticulated by touching the alveolar ridge with the blade or the front of the tongue instead of the tip of the tongue. This type of misarticulation is more common for learners with the previous knowledge of other Chinese languages.
ts'	(c)	s, ch	The nature of mispronunciation of this aspirated affricate is similar to that of /ts/ sound: Students do not compress air by firmly attaching the blade of the tongue to the upper teeth and the side rims of the tongue to the upper teeth. The closure /t/ is not formed and the closure release burst does not occur. Instead, a fricative /s / is pronounced with a slow air flow through the mouth. This error is usually distributed among beginning American learners. Another way of mispronunciation this sound is as an alveolar affricate /tʃ'/: The tip of the tongue is placed against lower teeth and the blade or the front against alveolar ridge. Learners with previous knowledge of a Chinese language often demonstrate this error.
tʂ	(zh)	tʃ	This sound is a retroflexed voiceless unaspirated affricate. The tip and the of the tongue is curled up and placed against the hard palate. The air is first compressed and then released with a slight aspiration. Students mispronounce this sound by simply raising the tip of tongue to the alveolar ridge without curling it up. The sound produced sounds "too front" while it is supposed to sound deep "back". Often this sound is mispronounced as voiced by beginning learners with the voice rising during the air release burst.

tʂ' (ch)	tʃ	This sound is a retroflexed aspirated blade-palatal affricate, with the place of articulation exactly the same as the unaspirated / tʂ /. Students mispronounce this sound by not curling the tip and of the tongue up, but simply touch with the tip of the tongue the alveolar ridge, producing an alveolo-palatal fricative. Sound produced with this articulation is too front and too soft instead of back. This misarticulation is common both American learners and for learners with previous knowledge of one of Chinese languages.
s (s)	ʃ	This sound is a voiceless dental fricative pronounced with the tip of the tongue resting on the upper teeth and air stream rushes through with a sibilant sound. The point of contact in the correct pronunciation is between the tongue and the upper teeth. The sound is piercing and penetrating. Students mispronounce this sound by bringing the blade of the tongue close to the alveolar ridge and without establishing a tight contact, letting the air flow through the mouth with strong friction. The piercing and penetrating sound is thus not produced.
ʂ (sh)	s	This sound is a retroflexed voiceless fricative pronounced by curling up the tip of the tongue and gently touching the alveolar ridge. The air stream rushes out between the tongue and the alveolar ridge thus causing friction. Students mispronounce this sound by either pressing the tongue too much on the alveolar bridge, causing a sibilant sound ; or misarticulate the position of the tongue by raising the tip of the tongue to the alveolar ridge. This mispronunciation pattern is more common for learners with the previous knowledge of one of the Chinese languages.
tɕ (j)	tʃ	This sound is an unaspirated voiceless palatal affricate produced with the tongue resting on the lower teeth, while the blade and the front of the tongue touch the alveolar ridge and the hard palate firmly. The side rims should be in close contact with the upper teeth , thus forming a complete closure in the mouth, with some air compressed in the closure. Then the closure is slowly released by gently lowering the tongue. In this way the air rushes trough the mouth making an unaspirated /t/ which immediately changes into / ɕ /. A common mispronunciation is caused by not firmly touching the alveolar ridge and upper teeth by the tongue, thus not forming complete closure in the mouth. As the result the air is not compressed and flowing through the mouth with strong friction. Sometimes this sound may also be voiced.
tɕ' (q)	tʃ	The tongue position and the movement are the same as for / tɕ /. The difference is a matter of aspiration, for /tɕ / it is slight and /tɕ'/ it is strong. The reason for misproduction is very similar to that for / tɕ /. Because the tongue does not firmly touch the alveolar ridge and the side rims of the tongue, the closure does not occur, the air flows through the mouth with friction during the whole period of sound production without release burst.

ɻ (r)	ɾ	This is a voiced retroflex sound which is pronounced by curling the tip and the blade of the tongue up to the alveolar ridge. Americans often confuse this sound with the English / ɾ /.
ɕ (x)	ʃ	This sound is produced by touching the lower teeth with the tip of the tongue, with the blade and the front of the tongue touching the alveolar ridge and the side rims of the tongue touching the upper teeth gently. The side rims should be in close contact with the upper teeth. The air stream rushes through the passage thus formed between the tongue and the hard palate and between the tongue and the alveolar ridge. This sound is misproduced by the tongue not touching the alveolar ridge and the hard palate close enough. As a result, the passage for the air is too broad and the noise produced is too strong, the sound resembles the fricative English sound /ʃ /.
p, t, k (b)(d)(g)	b, d, g	The most common mispronunciation is in producing these sounds clearly voiced.

Table 2. Segmental errors made by foreign learners of Chinese. (Finals)

y (ü)	iu	This sound is produced with the same tongue and teeth position as the sound / i / but with rounded lips. The Mandarin /y / is the same as the French sound /y /. Beginning learners often mispronounce this sound as the glide /iu/, which is a farther back sound, than the front /y / sound.
n (n)	ŋ	This nasal alveolar sound in the end of a syllable is often mispronounced as a velar nasal. The sound appears too back and may be confused with the /n/ sound distinctive pair which is the velar nasal.
ɤ (e)	ə	To produce this sound the back of the tongue must be raised and lips must be spread. The sound is mispronounced by raising the center instead of the back of the tongue and the sound similar to the English as in "third" is thus produced.
ɑw (ao)	aʊ	The vowel /ɑ / is misproduced as a front vowel, like in an English word "cow", while it is a back vowel.

From the review of the segmental errors made by students it was noticed that some patterns of mispronunciation significantly change the sounds, for example, /s/ misproduced as /ʃ/, or /ts/ as /s/. It can also be observed that mispronunciation patterns are distributed between two major groups of learners according to their L1: those who have English as L1 and those who have another Chinese language as a previously acquired language. We could compile another group of pronunciation error patterns characteristic for those whose L1 is Japanese or other European languages. For example, Japanese learners of Chinese have particular difficulty in producing rounded vowels, such as /u/, /y/ in such syllables as ***qu***, ***xue***, ***lu***; the / ***l*** / sound is also often substituted by /***r***/ by beginning Japanese learners. It was also reported that that Japanese learners of Chinese do not distinguish between / ***n*** / and / ***ŋ*** / and often substitute /***ŋ***/ with /***n***/. In addition to that, Japanese learners pronounce Chinese fricatives and retroflexes with significantly reduced friction. However, the learners with these language backgrounds were not encountered very often and the data base is not yet large enough.

The distinctive distribution of mispronunciation patterns among learners with different language backgrounds, as well as the significant change in the phonological appearance of segments in mispronunciation, confirm the observation made by Flege (1992: 590), that sounds of the L2 are differently perceived by learners than by L2 native speakers. The phonetic property of the L1 interferes with L2 sound perception and the gaps in the acoustic perception and analysis of new sounds in the L2 are being filled with the phonetic categories of L1 or their modifications. This implies that there is very little hope that learners would develop new phonetic categories merely from extensive spoken practice, without a systematic instruction and correctional feedback. Learners show that they do not perceptually distinguish the phonetic categories similar but not equivalent to the L1 phonetic categories, and thus can not accurately produce them unless they are systematically instructed.

Another observation made by Flege proposes that learners show better control over new sounds (for example, /*y*/sound for American learners of French), which are not associated with any phonetic categories from the L1, rather than similar sounds (for example, the English /*ɽ*/ sound for Spanish learners of English) which are associated with the phonetic categories in the L1 and thus cannot be learned adequately because of the strong interference of the L1 category. This has not been confirmed by my observing the learners of Chinese who had different language backgrounds. The mispronunciation errors are distributed rather evenly between the two groups of sounds. For example, fourth year American learners of Chinese often mispronounce "new" sounds: retroflexes as affricates and affricates as fricatives. At the same time they mix central and back position of vowels, do not always round lips when pronounce /u/ and /y/ showing misarticulation of the similar sounds.

4.2. A review of the method of teaching pronunciation to beginning learners of Chinese

The present model of teaching beginning Chinese in the Department of East Asian Languages and Literatures, of The Ohio State University is primarily focused in developing good spoken abilities, skills necessary for live and culturally appropriate communication. From the first lessons students are expected to perform short dialogues in Chinese at natural speed, with clear native-like intonation and stress patterns. The dialogues are introduced in "A Self-study Introduction to Beginning Mandarin Chinese", vol. 1-4 (Walker 1987). The dialogues and additional material in the workbook are recorded on a set of audio tapes. To meet the requirements for performance, learners have to spend most of their preparation time working with audio tapes, learning by imitating the model voice on the tapes and following the guidance on articulation positions for accurate producing segments. In addition to introducing the dialogue, the tapes also have extensive sets of drills from which students can learn to pronounce individual words and patterns used in the dialogue and other material covered by the lesson. Though the tapes do not drill the segments of Chinese separately from the words, students still have the auditory input of the L2, which is used as a model for imitation, and which gives a hint as to how Chinese sounds are pronounced by native speakers.

The structure of Chinese syllables and individual sounds is more completely described in the "Pronunciation and Romanization Module" (P&R) (Interagency Language Roundtable 1979). This module is combined with a set of six tapes and six computerized diagnostic tests. Reading this book and listening to the tapes, students can gain deeper understanding of the structure of the Chinese syllable, learn the existing syllables in Chinese and get familiarized with Chinese tones. This workbook gives explanations as to how to pronounce the sounds of Chinese in terms of the place and manner of articulation and suggests how to avoid possible mispronunciation errors. A set of exercises for training pronunciation overtly with the tape reinforces theoretical knowledge about Chinese sounds by practical application. The exercises are built on the principle of contrasting pairs: aspirated-unaspirated, rounded-unrounded, closure-fricative; and on the principle of gradually changing the tongue position: palatal, alveolar, dental; thus stimulating the development of articulation movements.

The "P&R" uses *pinyin* as a written representation, for "transcribing" the sounds of Chinese. Beginning students learn to distinguish Chinese sounds in terms of the written *pinyin* symbols which help them identify sounds when pronounced unclear or mispronounced and to demonstrate their ability to distinguish Chinese sounds as they perceive them.

The "P&R" is divided into six stages, which contain practical exercises on pronunciation learning also recorded on tapes. The "P&R" exercises are to be done from the tapes and from supporting materials in the book. Each of the six stages in the "P&R" are implemented in the first unit of the workbook and are assigned to be done accordingly. The student's progress within the "P&R " program is tested on the acquired ability to recognize Chinese sounds and tones from the model voice on the tape or from the computer and to be prepared to write them down in *pinyin*.

The six computerized diagnostic tests correspond to each of the stages in the "P&R". The tests were designed especially for enabling both learning and testing of listening recognition skills for Chinese sounds and tones as well as the *pinyin* Romanization. The results of the tests are evaluated according to the percent of correct answers. The evaluation of diagnostic tests also presents detailed reports on error analysis. From these reports students can know exactly what error was made and which response should have been given instead. A student may play back the sound pattern which was reported as misunderstood and analyze the error. The diagnostic tests, together with the "P&R" and tapes, create a solid basis for learning to recognize Chinese segments and tones. However, the ability to pronounce the sounds and the sound patterns introduced in each set is not tested. The exercises on training pronunciation in the module, which learners fulfill as a part of their preparation for classroom activities are not tested accordingly. The "P&R" homework on learning sound and tone production is considered to assist in performing dialogue and supplemental vocabulary material in class, and is thus tested during regular classroom activities. The sounds and patterns suggested for pronunciation practice in the "P&R" do not exactly match the sounds and patterns introduced in the dialogue and vocabulary in the referred stages in the "Self-Study Introduction" workbook. For example, in the first stage of the Workbook such words as *xiansheng*, "Mr., term of address"; *xiaojie*, "Ms., term of address"; *tongzhi*, "comrade"; *laoshi*, "teacher"; *xuesheng,* "student" are introduced for learning, while the referred part in the "P&R" introduces Chinese tones through practicing them in such syllables as *fang, wei, mi, wu, yin.* In the same "P&R" set consonants and vowels are introduced for practice: *ma, wu, meng, an, en, yin, hong; han, liang, miao, nie, fang*. Guidance as to how to articulate the new sounds: *x-*, *sh-*, *j-*, *.zh-*, *sh-* is introduced on the audio tape and students learn and practice articulation positions along with memorizing dialogue and vocabulary items on the tape. When words with new sounds are introduced on the tapes, the articulation positions for these sounds are explained. This explanation is very important because it allows students to start developing new articulation positions and to realize the differences in the L1 and L2 articulation patterns. From the moment when the obvious difference in articulation of new sounds of the L2 has been realized, students start developing a new phonological structure in memory distinguished from that of the L1. If the expanded explanation of sounds' articulation would be also placed in the "P&R", illustrated with schemes of the articulatory organs during the pronunciation, students would obtain better support for understanding articulatory mechanisms. It would also make this information more accessible, because many students only listen to the tapes in language labs but have access to the books at any time.

During dialogue memorizing and performance, students are primarily focused on learning to accurately pronounce short phrases of connected meaningful speech. Students are supposed to perform the dialogue as close to native manner as they can. Usually the dialogues are simple in terms of grammar and are not abundant in new vocabulary, thus students' attention is not strongly distracted from the primary goal of mastering pronunciation skills.

Along with learning pronunciation, students learn how to produce words and patterns in the meaningful interaction environment created in the classroom and within a culturally appropriate context. This helps to associate the auditory (phonological) and visual (written) information with the semantic meaning which is important in learning speech sounds as parts of meaningful units: words and sentences. The association of sounds with meaning in memory is the most powerful vehicle for transferring the L2 phonological information into LTM (Chapter II: 2.5.2; Chapter III: 3.4) .

The performance of the memorized dialogue usually starts classroom activities and is followed by the set of exercises. The components of the dialogue are corrected after the performance when students are instructed to rehearse the difficult vocabulary items from the dialogue, primarily aiming at improving pronunciation. As a rule, rehearsal is conducted with the whole class, in order to involve everyone in sound learning and is followed by a short time individual sound production. The large number of students in a class does not allow a sufficient amount of time for training individual pronunciation drills. Individual error analysis and pronunciation feedback is impaired.

During the exercises which follow the dialogue performance, the instructor involves students in group interaction, in which students extensively use the vocabulary and patterns learned during home preparation to communicate with the instructor and with each other. The context of the exercises gives students the opportunity to choose a correct response among the set of possible responses and to show their understanding of the semantic meaning and cultural appropriateness of the context. If a student's response is reasonably prompt, semantically correct and appropriate culturally, then the student is usually encouraged, even if his/her pronunciation could be more accurate. There are no exercises, besides dialogue performance, focused specifically on improving pronunciation skills in class. Consequently, students do not receive enough systematic feedback about their pronunciation. The correction they do receive is usually superficial with errors being corrected through multiple repetition until the sounds produced by a learner have been adjusted to native-like norms. Deep pronunciation error correction, which implies an analysis of the reasons for the sound misproduction and an explanation of correct articulation, is usually not done.

We can notice two main factors that influence learning pronunciation at the beginning level: first, the large number of students in class which impairs individual error analysis and correction feedback. Second, the classroom exercises focus on primary achieving fluency in performing an ever increasing inventory of linguistic items and capacities within meaningful and culturally appropriate communication (Walker 1989:54). Pronunciation is considered to be improving gradually along with extensive practice of communication skills and due to home practice with “P&R” tapes and exercises.

As a result of this method of teaching pronunciation, learners by the end of the first year develop good listening comprehension ability and reasonably fluent production of mostly pedagogically presented speech within culturally appropriate context, but demonstrate interference of their L1 with the pronunciation of sounds and tones.

The difficulty of acquiring native-like Chinese pronunciation through extensive communication practice can be explained by Flege’s hypothesis that non-native speakers not only do pronounce sounds of the L2 using the L1 phonetic categories, but also perceive them in terms of the L1 phonological categories (Flege 1992: 568). The two types of memory, declarative for listening recognition and procedural for sound production, have their particular properties and require different types of input for their development. It was observed that although the two types of memory do not match in certain steps of their development, (for example: good listening recognition skills and poor production skills), they mutually affect each other. Poor or strongly accented production of L2 sounds by a learner affects the way L2 sounds are stored and may result in preventing the L2 phonological information from being stored authentically. Sound production can be considered as an additional L2 auditory input which interferes with the L2 authentic auditory input. We can assume that, even if authentic auditory input of L2 sounds was received and stored, it is strongly affected by the learner's accented L2 production. Thus the L2 acoustic output must be accompanied by a sufficient and careful error analysis and correction feedback in order to provide an accurate input which would reinforce authentically stored L2 sounds in memory.

Let us review the models of learning L2 pronunciation proposed in Chapter III in terms of their correspondence to the actual methods of learning pronunciation practiced at The Ohio State University. This review will help to explore the model of learning L2 pronunciation introduced in Chapter III and what this model can propose for the improvement of the present state of teaching pronunciation in class.

Step one in the model is to learn to distinguish and to recognize the sounds of Chinese from auditory input. A set of tapes which goes with the "P&R" and the "Self-study Introduction" workbook provides the necessary auditory input for starting to develop the L2 phonological structure. How well a student has learned to recognize sounds auditorily can be tested through their ability to "transcribe" them in written form. Fro this purpose *pinyin* is introduced and employed as a means for testing students' listening recognition ability . Each student has to pass six computerized diagnostic tests which comprehensively revise the listening recognition ability of students. As far as most students, as a rule, show good results when passing listening recognition tests, we can conclude that the auditory input was sufficient, was supported by the necessary theoretical explanations about the distinguished features of the L2 phonetic categories, and sounds have been learned to be identified in written form by means of *pinyin. Step two* in the model of learning L2 phonological structure (Fig. 4) is, therefore, introduced in the program of teaching pronunciation in the Chinese language program practiced at The Ohio State University. The order of introduction the information -- first auditory, then visual also agrees with the model of learning L2 phonological structure.

Step three in the model of learning the L2 phonological structure is attaching meaning to sounds and learning L2 phonology in a meaningful context. The memorization of dialogues and participation in meaningful interaction during classroom activities develops the ability to recognize meaning in combinations of sounds. This step enables students to develop listening comprehension of connected speech, which is an upper level operation in sound recognition. Students develop the ability of *conceptually driven* or top-down processing of speech, which means that the language command and the subject matter affect recognition, in contrast with the *data-driven* or bottom-up processing, which is affected by the ability to recognize particular sounds and depends upon the clarity of sound production. The two types of processing, however, interact in order to eventually determine listening comprehension and both are important for learning L2 sounds.

For testing students' listening recognition and comprehension of phonological information in a meaningful context, the audio recorded set of "Comprehension dialogues" was implemented into the program. Each set contains from five to eight short script-driven and authentic dialogues pronounced by native speakers and preserving the attributes of spoken interaction of Chinese. Students have to listen carefully to the dialogue and answer the questions of the instructor in order to demonstrate their ability to comprehend the auditorily perceived speech. Dialogues are pronounced at a natural speed and use some unknown vocabulary. Students are expected to extract the basic information utilizing the knowledge they have obtained so far. The step three in the model of learning pronunciation which is analysis of auditory information through its meaning, therefore, is also employed in the present method of teaching pronunciation.

S*tep four* in the model of learning L2 phonological structure is to acquire the skill to reproduce sounds silently, e.g., to retrieve from memory previously stored authentic auditory information and to hear it over again, without pronouncing aloud. This step in learning pronunciation reinforces the auditory information stored in the memory without affecting it with the inaccurate input from the learner's production. Before performing the dialogue in class, one can see how students rehearse the dialogue by "mumbling" it in a very low voice or whispering the text after the audio tape. This performance cannot be equated with what we call "silent re-production". When mumbling or whispering the phrases, students hear what they produce and not what they initially heard from the tape. Students do not learn to retrieve sounds in memory and to hear it before sounds would be pronounced overtly.

As it has been already mentioned above, production strongly affects how L2 sounds are stored in memory. When sounds are persistently misproduced by the learner, L2 sounds cannot be authentically stored. When at the level of learning the production of L2 sounds, it is important to develop in learners the ability to notice the difference between what they wanted to produce and what they actually produced, e.g., to self-correct their mispronunciation errors. This ability can only be mastered under persistent correctional feedback. A complex of exercises recorded on audio and video tapes designed especially for pronunciation practice must find its particular place and time in the program of teaching the L2. Each student would be able to listen, produce, listen to the model over again and to his/her own production and then self-correct the performance.

Learning pronunciation must not be overlooked at any stage of language learning in order to develop stability of pronunciation skills. Practice has shown, that in the first level of L2 instruction (Walker 1989: 49) students achieve good initial pronunciation skills because pronunciation is attended to by the instructor. On the upper levels, when pronunciation is less attended, learners do not have sufficient feedback and thus their pronunciation skills are behind their general spoken abilities. For example, dividing long sentences into short meaningful units with pauses, and correct stress, tone and intonation in connected speech needs a special learning program for developing accuracy in pronunciation skills along with interactional abilities.

Having analyzed the program of Chinese which is being practiced in the Department of East Asian Languages and Literatures and how it corresponds to the proposed models of learning L2 pronunciation, we can conclude that the strategy of the program generally agrees with the models, although in certain tactical issues both the models and the program could be mutually enriched. Being from the very first steps focused on participating in meaningful interaction with an extensive use of L2 patterns and sentences, students rush through a very important step which must not be overlooked. This step is in learning to adequately perceive and accurately produce Chinese phonetic categories and in the ability to differentiate L2 sounds from L1 sounds. Learning to distinguish the L2 sounds from the L1 sounds would enable students to develop new phonetic categories in memory and is a necessary precondition for not developing a strong accent. At this point the learner's ability to distinguish the L2 phonetic categories can be tested by their ability to explain the manner of articulation for the sounds of L2 and the difference between L1 and L2 familiar and new sounds. When taught to produce sound categories accurately, students can be insured from developing inaccurate phonetic categories in memory which would inevitably affect pronunciation. Thus a more sufficient input for learning segmental structure of the L2 phonology at the beginning level is necessary. Instruction of segmental structure in listening perception is fairly well developed in this program (series of tapes for the 'P&R', "A Self-study Introduction"), but exercises for training L2 segment production in the classroom and at home could be developed and implemented on a more systematic basis. Correctional feedback needs to be provided in order to make these exercises effective. Experience shows how effectively pronunciation skills can be improved with the help of exercises particularly designed for training pronunciation. When along with learning pronunciation, students during the exercises also have to think about grammar and semantic aspects of the vocabulary, the focus on pronunciation is significantly impaired. This means that a certain amount of time in during the course instruction should be spent for the specific purpose of pronunciation learning in the same way that a certain time is spent for presenting grammar, commenting on vocabulary and cultural context. Pronunciation should not be considered as a skill that will develop automatically by increasing the load of production experience, without a guidance.

Observations of experienced learners who have lived in the country of the target language for five years and developed native-like vocabulary, grammar and knowledge of cultural issues but didn't sufficiently improve pronunciation using the L1 phonetic categories to substitute the L2 phonetic categories, show that pronunciation can not be mastered by practical application only, without a sufficient guidance.

Another observation is that students can either lose or improve their pronunciation skills on each level of language learning depending on how carefully pronunciation issues are attended in class. For example, if pronunciation is focused on suprasegmenal features and segmental features are left unattended for some period of time, then most probably students would switch to L1 phonetic categories without noticing it because they are still easier to pronounce. Observations at student's behavior on different levels of L2 learning confirms that pronunciation learning is a long-term process and that the guidance correctional feedback is necessary at every level of instruction.

4.3. Guidelines for teaching Chinese pronunciation proposed for the four-level Intensive Chinese Curriculum Design

In the four-level Intensive Chinese Curriculum (Walker 1989), a framework for teaching basic skills for language acquisition was proposed. The courses in four skill areas: spoken development, spoken application, reading and writing were conditionally divided into four levels according to the amount of classes taken and the level of acquired skills, rather than the traditional division of levels according to academic year or any other chronological unit. Developing the guidelines for teaching pronunciation, we used the four level framework for Spoken Application courses as a basic program. We assumed that the program of teaching Chinese pronunciation based on the proposed guidelines would correspond with the goal of the course aimed at the general improvement of the spoken abilities of learners.

Level one. Instruction should be primarily aimed at establishing L2 phonological categories in the memory of a learner and develop automaticity in accurate production of segmental features of L2. Students have to learn to distinguish L2 phonetic features perceptually (listening recognition) and must be able to explain the sounds' differences in terms of the place and the manner of articulation. The theoretical knowledge of the articulation positions for L2 learned from workbooks and/or from instructor gives students confidence that his/her practical articulation is accurate. L2 phonetic categories are divided into “new” and “similar”. New L2 sounds are more quickly attended to by learners and the differences with the L1 are more easily realized, which facilitates developing new phonetic categories in memory. Similar L2 sounds are easily confused with L1 phonetic categories and are easily substituted by them when the actual phonetic differences are left unattended (Chapter III: 3.2.3). Not establishing new phonetic categories for similar sounds in the memory prevents developing the new phonological structure, but it leads to adding some new phonetic categories to L1 phonological structure in the memory. This results in a strong tendency to never develop accurate production of L2 sounds. Strong interference of L1, therefore, will become a part of speech production.

The establishment of new phonological structure on the level of perception must be accompanied by developing new categories on the level of sound production. The two modalities of memory connected with perception and production of sounds strongly influence each other. The strong interference of L1 in sound production of L2 affects the way the L2 sounds are stored in memory. Sounds cannot be stored authentically in the memory of a learner if they are inaccurately pronounced. Conversely, when L2 sounds are not perceived authentically, for example, by auditory input and theoretical explanation of the articulatory features, they cannot be stored authentically, as new phonetic categories, and cannot be learned to be accurately produced. (Fig. 5) Therefore, on level one of learning pronunciation of Chinese sounds, the students’ attention must be concentrated on two issues: getting sufficient auditory and theoretical information about the new and similar Chinese phonetic categories in order to store them authentically and, second, to develop automaticity in accurate production of segments.

On suprasegmental level learners whose L1 is not a tonal language have special difficulties in learning Chinese tones. However, once established, the tones are not affected by the tones from L1, as it happens when learners L1 is a tonal language, but are affected by the stress system and intonation in their L1. Listening recognition and production of tones are different memory capacities which nevertheless are mutually dependent. In order to pronounce tones correctly, a sufficient authentic auditory input must be provided. Theoretical explanations about how voice pitch must be regulated while producing different tones must support the auditory input. Tones are easier to establish when they are first pronounced in isolated syllables. Following the linear predictability constraint proposed in the delta rule of learning, the next step should be learning to pronounce tones in one, two and three syllable words, and then in sentence production. A set of special exercises must be provided in order to develop automaticity in producing tones. Special attention in these exercises must be drawn to the tonal behavior in stressed and unstressed syllables and in tone sandhi. Computer assisted programs that visually reproduce the contour of the voice pitch in different tone production, can be very helpful in learning tone production.

Tones in most cases are bound to a word's meaning in Chinese (unless a syllable can be realized with only one tone, as in /néng/). Introducing tones with the meaning of the word makes learning tones easier because when associated with meaning they establish links with the semantic information stored in LTM and are transferred to the long term memory (Chapter II: 2.5.2).

To conclude the proposed requirements for learning pronunciation on level one, we can say that the primary goal is to develop automaticity in accurate production of segments and tones. During the performance of short, pedagogically designed dialogue, students should be able to demonstrate clarity of sounds and tone production in short sentences.

Level two. Spoken Application Class in level two is in general focused on "practicing making inquiries and coping with spoken Chinese that is not pedagogically controlled" (Walker 1989:54). Unlike level one, where performance is basically script oriented, and extracurricular conversational exchanges with instructor are limited, level two leads students to developing skills for using language in context oriented instruction. The communicative exchange is moving from automatic participation in formulaic exchanges to interaction that is aimed at extraction of information from instructor with the help of making inquiries in Chinese. Guesswork is a new strategy introduced on this level. Non-rehearsed communicative tasks based on mostly non-pedagogically designed material, lets students make their first independent steps in communicating in Chinese. It is important to assist learners to develop the feeling of comfortable participation in non-scripted communication and to move forward in the language. On this level, most of learners' attention is concentrated on the meaning and grammatical correctness of what they say. In order to develop accuracy in the new form of language application, learners must be equipped with the skills of self-control over his/her pronunciation. Now when an automaticity in accurate segment production has been trained, much attention can be given to the tones in connected speech and intonation patterns. If most of the time in spoken application class is spent on training the ability to ask and answer questions, then it is a good chance to concentrate on developing automaticity in accurately pronouncing interrogative and declarative patterns. The voice pitch contour for each sentence type must be explained and the difference between the English and Chinese pitch height should be highlighted. In order to provide sufficient auditory input for interrogative and narrative sentences, it would be helpful to provide tapes with minimal distinguished pairs on the basis of very simple Chinese sentences first. For example, simple construction like the following:

Tā něi yì nián lái Běijīng gōngzuò? --- *Tā qùnián lái Běijīng gōngzuò.*
Which year did she come to Beijing to work? - She came last year to work to Beijing.

can be used for exemplifying the difference in pitch height. By learning to first realize the difference between the two sentences in a pair students more likely are prevented from engaging in trial and error activities that need to be corrected. For drills outside class, a speech analysis program can be used. These programs can display a pitch height and a pitch contour, enabling visual feedback of tones and intonation patterns. Speech analysis programs can provide necessary correctional feedback which can save a lot of drilling time in class.

In addition to correct pitch height and pitch contours in connected speech, accurate phrase production is also characterized by the ability to divide phrases into complete meaningful units making appropriate pauses and placing stress in the right location. Students' progress in pronunciation must be systematically evaluated (every two weeks). Performances in the form of presentation or regular class sessions should be videotaped and analyzed after class by both instructor and learner. It would be better if the instructor can do it in a quiet environment concentrating on one learner's performance at a time. The instructor should make comprehensive notes about the learner's production. Instructors usually do not have an opportunity to conduct a deep error analysis of learners' pronunciation in class, because his/her attention is concentrated on more noticeable errors, such as grammar and vocabulary. Instructors are also very busy with leading classroom activities. In addition to that, students sometimes feel uncomfortable when their performances are corrected and discussed in front of the class, especially if they feel that they are corrected more often than others. An effective way of correcting pronunciation is when instructor simply repeats after the student his/her phrase but with accurate pronunciation, letting all other students hear it, but without pointing to the mispronounced part. This way of correction lets all students hear the correct version without detracting their attention from the focus of the classroom activities. Students may be assigned for individual appointments during which they can analyze their performance recorded on the tape together with their instructor. Students can be invited to discover and identify his/her mispronunciation errors first, while the instructor suggests his/her version of what needs to be corrected afterwards. The instructor can also prepare an observation list where mispronunciation, as well as grammar and other errors would be included together with his/her suggestions for effective error correction. Such individual sessions may have positive impact on different aspects of performance and lists of mispronunciation errors could create a valuable data base of the most common mispronunciations made by students on level two. In the future this data base would become a valuable source of material for developing special exercises that would allow to deal with the problem of mispronunciation on more systematic basis.

The focus of the control over the pronunciation on level two moves from immediate, drill-oriented and mostly in-class activities towards more individualized, analytical and after-class activities. On level two, when learners are supposed to move from script-driven to context-driven performance and start to exploit their source of the language as a means of the non-script driven interactions, they should be encouraged to make new steps forward and to gain new experience. That is why making explicit error correction in class may be psychologically harmful for students. In the new classroom environment the student's control over learning pronunciation is significantly impaired and needs a reinforcement of systematic correctional feedback, which will be more effective when conducted in the after-class sessions. Comprehensive and systematic control over pronunciation from the instructor along with exercises on computer-based instructional program on training pronunciation should be provided. Persistent correctional feedback from the instructor at level two is important because the students' L2 phonological representation has not yet stabilized. When receiving inaccurate auditory information about the L2 phonological elements from his/her own inaccurate pronunciation, students shift the previously stored L2 authentic phonological information towards inaccurate and gradually lose previously learned knowledge. Deliberate grading of progress in pronunciation would encourage students to develop skills for the self-control.

Level three. By level three students must develop self-control over basic pronunciation skills: production of segments, tones, stress. The differences in pitch contour and height in interrogative and short declarative sentences must be acquired. Self-control over pronunciation means that a student got the ability to self-guided learning (Bechtel and Abrahamsen: 74). This implies that the learner has developed a stable L2 phonological structure in memory which enables receiving the auditory input from outside (from other sources or the output from the learner him/herself) and comparing the input to the model structure in memory. The discrepancy between desired and actual input will be evaluated and the ways of reducing the discrepancy between the two will be suggested, thus providing the process of self-correction and learning. On levels one and two of learning pronunciation auditory and visual capacities are predominantly employed for processing and storing phonological information. Accordingly, auditory and visual stimuli are most often used as retrieval cues. In monolingual classes students are involved into series of question-answer interactions where they are given auditory or visual information as retrieval cue. On level three the semantic capacity of memory is going to be primarily employed due to the tasks that students are assigned to accomplish. According to the Intensive Chinese Curriculum (Walker 1989), level three course is focused on developing skills of sustained narration in L2. On the basis of video, script and audio of the sound track of a feature film students are supposed to create narrations and present them in front of the class and instructors.

Students do not participate in the first and second person exchanges as much as in the previous levels, but play a role of a third person narrator which is a new experience for them. Meaning of words rather than phonological encoding becomes the most widely used retrieval cue on this level. It presents certain challenge to students because the narrative performance employs a different function in memory -- *recall* -- instead of the *recognition* which is broadly used in question-answers exchanges in lower levels. (Chapter II: 2.6) If in the levels one and two students could elicit most of the necessary information for their response from the context which provides *recognition* cues, in sustained narration students are supposed to demonstrate how well they can *recall* phonological representation of the meaning. On this level one could often observe a "tip of the tongue" phenomena, when a student had the feeling that he knew the necessary word , but could not recall it at that moment. It means that different cues were previously used for retrieving this word and the semantic cue did not have an effective retrieval path. The inability to accomplish retrieval operation will be reflected in long pauses in inappropriate parts of a sentence, broken patterns and substitution of the searched word with other words or phrases. This way of presentation significantly impairs listener's understanding and lowers the accuracy of the speech production on the whole. What can the model of memory propose to help learners in this situation? According to the Distributed Model of Human Learning and Memory, each module in memory combines a number of different sources of information. Some input is received from sensory-oriented modules, others will be received from abstract modules, which, in turn, receive input and send output to other abstract-oriented modules of different modalities of memory

(McClelland and Rumelhart: 175). Therefore, abstract-oriented modules receive input both from abstract modules and from sensory modules. By the beginning of level three, students used to employ mostly sensory (auditory and visual) cues for retrieving phonological information. At the level three sensory retrieval cues are employed along with abstract (semantic) retrieval cue. If the auditory and visual information about the text to be narrated will be employed during the text remembering, then the student would have more retrieval paths which would support semantic retrieval cue. For example, a text is divided into paragraphs. Each paragraph requires a special intonation, rhythm and stress. In addition to that, in Chinese a tonal melody of a sentence is a salient feature for its remembering and retrieval. When remembered and narrated later in class, the knowledge of the tonal picture of the phrase can become a strong cue for their reconstruction. This way of remembering a text employs the mnemonic technique used for memorizing poems or short dialogues, familiar to students from the previous levels. Visual mnemonic technique also can be employed for reinforcing retrieving. Many students remember how words and phrases are located on a page and this helps them to recall necessary words during speech acts. New or particular difficult words can also be placed on margins using a different color of pen. In Chinese the form of a character or characters which imprint in visual memory during memorizing a text can be a powerful visual retrieval cue.

Level four. By this level learners must establish good control over retrieving the sensory (auditory, visual) information with semantic cues. The connections in memory responsible for associating different types of storing phonological information have been reinforced. Thus, by the beginning of level four students are expected to have stabilized representation of Mandarin Chinese phonological structure in memory. In terms of production abilities, learners develop skills to accurately pronounce L2 segments and native-like intonation patterns in prepared speech. It supports the assumption proposed in the model of learning L2 pronunciation (Chapter 2) that after learner has equally developed three types of representation (auditory, visual and semantic) of the L2 speech sounds in memory, he/she can feel confident in using the phonological properties of the L2 within a prepared context. In addition to that, the learner builds a solid basis for future independent learning of Chinese phonology. What leads us to think that learners with the degree of acquisition achieved by the level four is able to learn and improve his knowledge of Chinese phonological system in the future without an outside guidance? We again refer to the Distributed Model of Human Learning and Memory. The mechanism of learning in the model proposes that the weights of interconnections between activated units in memory are self-adjustable to the external input due to the ability to drive the weights to right values to make the internal inputs to the unit match the external inputs. The prerequisite to achieving the self-adjustable state is establishing a well-developed and functionally effective structure. This structure must have a ramified network of interconnections among units in memory which are able to perceive and get activated by diverse external input. If the system of interconnections is not sufficiently developed and ramified to perceive various input, then the network would not be able to respond to the unfamiliar input i. e. there would not be no particular units that would be activated by this input and thus start the process of adjusting weights (linear predictability constraint principle). Another precondition for enabling the process of self-adjustment of weights of interconnections among units is that the external input is significant enough and gives sufficient excitatory input to activate the units in memory. Thus, for achieving self-learning state the structure in memory must be stabilized and ramified by numerous interconnections with other units. In addition to that the external input must be sufficient to be able to activate units in memory.

By level four students develop central representations for segmental and suprasegmental features of Mandarin Chinese. It means that a learner can extract a central tendency in a variety of productions and distinguish accurate production form occasional misproductions by other learners. The Distributed Model of Human Learning and Memory calls this "the prototype of the central tendency" principle and the ability to extract the prototype from occasional distortions "the key aspect of the model's behavior" (McClelland and Rumelhart: 182). According to Flege, the ability to extract prototype in L2 phonological representations comes with extensive expansion into the language speaking environment which gives a learner the opportunity to compare and to extract prototypes (Flege. 1992: 593). This ability can be developed in a shorter period of time if a learner has been guided to distinguish varieties of authentic pronunciation from mispronunciations. If over time a learner has not been guided to extract the central tendency in pronunciation forms then, according to the model, the learner would have to adjust the previously stored phonological representations to the new external input, including mispronunciations from other learners.

According to the Intensive Chinese Curriculum, a Spoken Application Course on level four is focused on the performances that a foreigner participating in Chinese culture is likely to find important in fulfilling his or her intentions and obligations (Walker 1989:59). What can this level suggest to learners in terms of enrichment of their knowledge of Chinese pronunciation? To my experience of functioning in different areas of China where Mandarin Chinese is considered the main language of communication the greatest problem that a learner comes across with is the different regional varieties within Mandarin norms of pronunciation. Sometimes varieties are so significant that it takes several days to familiarize yourself with the patterns of the varieties. It is easier to achieve understanding when the topic of communication is familiar to you, and when the topic is not known, then it takes even longer before complete understanding will be established. The misunderstanding goes only in one way and not the other: Native speakers usually do not have problems with understanding the standard pronunciation norms which are used by foreigners. I think that a learner would be more successful if he/she would be prepared to meet with the local varieties of pronunciation beforehand, for example, by listening to the tapes which recorded local news broadcasting from different areas where Mandarin is considered a norm of pronunciation but the varieties of pronunciation are remote from standard norms. Chinese students from other departments of The Ohio State University who came from different parts of China and can speak Mandarin can be invited to talk to learners or give an interview on one of the topics included into the syllabus. This meeting should be recorded on audio tape and during the following session the varieties of pronunciation be reviewed and analyzed with learners of Chinese. This listening comprehension practice would allow learners to expand their knowledge about the Mandarin Chinese phonological representations without interfering with the standard norms learned as central representations. This phenomena of coexistence in memory different varieties around one central representation was relied upon by the Distributed Model of Human Learning and Memory (McClelland and Rumelhart: 182): The model proposed that representations of specific, repeated exemplars can coexist in the same set of connections with knowledge of the prototype. The more specific representations a student can store the easier it will be for him to adjust to the environment where regional varieties of pronunciation differ from standard norms.

In terms of pronunciation skills on level four, as a result of quitting memorization practice from tapes students have a strong tendency to pronounce Chinese text with a distinct English intonation. It is especially common during their presentations in front of the class. While making their best to deliver the contents of their presentation and concentrating on the grammar and vocabulary issues, students sometimes completely switch to their L1 intonation. The problem of stabilizing the production of Chinese intonation can be helped through the training exercises that are recommended to TV and radio broadcasters: memorizing short messages produced by models with the following performance of them with the authentic intonation and pauses and at a normal broadcasting speed. Students can choose to memorize the parts of weather, sports, health or local news broadcasters from the Chinese TV daily news program and then to play the "News program" performance in class or during speech contest.

4.4. Summary

In this chapter we proposed guiding principles for teaching Chinese pronunciation within the four level Intensive Chinese Curriculum Design. When proposing the guiding principles we relied upon the models of learning L2 pronunciation discussed in Chapter Three and the analysis of the most common mispronunciations made by learners at different levels. In order to better realize the reasons causing the development of the accented production, we reviewed the existing method of teaching pronunciation from the point of view of the models of learning L2 pronunciation into the program. We observed that the three steps of learning: Analysis of auditory, visual and semantic information were employed in the program. The fourth step which requires silent production of the analyzed phonological information, however, was not emphasized and practiced and but was substituted by overt vocalizing. The reason, that learners misproduced sounds, despite the developed program of teaching pronunciation, was proposed to be in the insufficient error analysis and correctional feedback. The learners usually receive only a surface error analysis that allows them to adjust their production to native-like but not to acquire new articulation patterns. The exercises and drills for acquisition of the new pronunciation patterns are designed for students' practice outside of class with the tape. The pronunciation exercises are not integrated into classroom activities. Thus, the correctional feedback is impaired and error analysis does not follow mispronunciation.

The guidelines proposed in this chapter are aimed at emphasizing pronunciation learning within the Spoken Application Course. The recommendations for pronunciation learning were provided for each of the four levels of instruction. These guidelines can become the foundation for the future program of teaching pronunciation in the Intensive Chinese Curriculum Design.

Chapter V. Learning and teaching Mandarin pronunciation from the perspective of memory construction.

This thesis investigates the process of learning the phonology of L2 from the point of view of the psychology of memory and learning. It investigates the possibilities of acquisition of native-like pronunciation of L2 by adult learners and the factors that affect the mechanisms of learning pronunciation.

In order to understand better the process of learning L2 phonology, we took a broader look at the existing psychological models of learning and memory. The Parallel Distributed Model of Human Learning and Memory gave us a clearer view of the processes in memory which enable learning. Although this model was designed for application in the field of psychology and not pedagogy, its value for us is in that it has summarized the previous knowledge of memory and learning acquired in different areas of memory research in psychology and neuropsychology. The basic principle of the model, called the delta rule of learning, explains the process of learning as the adjustment of weights in the interconnections of units activated by an external input in order to match the output to the input. The principle difference of this learning rule was in that it explained memory as a distributed model which was "capable of accounting for empirical data that has been taken as suggesting that we store summary representations (e.g., prototypes) as well as data that has been taken as suggesting that memory consists of an enumeration of specific experiences" (McClelland and Rumelhart:17). The model has explicated storage of both general and specific information. In comparison with the previously used models of memory which claimed that memory is either a store of generalized representations i.e., abstracted representations of concepts discarding specific features; or prototype -- representations of typical exemplars (Rosch 1975), or a store of enumeration of specific experiences i.e., each word or the representational object is itself the result of a conspiracy of the entire ensemble of memory traces of the different individual experiences we have had with this object Jacoby (1983a,b); Hintzman (1983); Whittlesea (1983). The Distributed Model of Human Learning and Memory, therefore, is a possible key to our understanding of issues in the learning processes; for example, how we can store at the same time both the central representation of a sound and its allophones without confusing them and why the familiar sounds of L2 are more prone to mispronunciation than the sounds that are dissimilar to the sounds of L1.

Another extremely important implication of the delta rule of learning is that learning is categorized into supervised and non-supervised, and that these are two different types of learning which produce different output. In the supervised model of memory and learning, the activated network is informed which output is desired and the actual output is gradually adjusted to what is desired. The learning proceeds according to the bias of supervised learning. In non-supervised learning, the desired output does not inform the network and the adjustment of weights of interconnections of units may go in any direction. We took this observation as a basis for our assumption that in supervised learning, pronunciation of the L2 can be gradually adjusted to the desired output, e. g., native-like pronunciation can be achieved if sufficient feedback is provided.

This thesis proposed models of learning L2 phonological structure and of learning sound production. Different modalities of memory: auditory, visual and semantic, are involved in perceiving, analyzing, and storing of L2 phonological information. Although the three modalities of memory are simultaneously involved in the analysis and storing of phonological information, the L2 phonological information must be presented in the order of auditory - visual - semantic for most effective learning. Focusing on one type of representation of the information reinforces the corresponding modality (for example, written representation of information corresponds to the visual modality in memory). This impairs involvement of other modalities in analyzing and storing. Therefore, the phonological information must be presented in the order proposed in our model, for example, written representation and the meaning should not be presented before auditory representation, and semantic representation should go after auditory and written representations. This model organization was supported by Hector Hammerly (1991:121):

> "Young children acquiring their native language unconsciously through communicative interaction may also proceed globally, in a similar semantically focused word- and sentence- generation order. Their attention is almost exclusively on meaning and they can develop lower-level distinctions at leisure. Adult learners must necessarily proceed in the opposite direction, with step-by-step guidance, if they are not to internalize fault rules and elements. In classroom SL (e.g., L2) learning for good results the focus cannot be at first primarily on meaning. Through the intermediate level it must be primarily on form for otherwise students do no master lower-level distinctions (involving sounds, morphemes, and certain syntactic rules) and develop deficient linguistic habits".

These principles are integrated into the model of learning L2 phonological structure. The analysis of the processes of analyzing and storing L2 phonetic and phonemic L2 information in memory, we concluded that adult learners are under strong cross-linguistic influences which prevent them from perceiving and categorizing L2 sounds authentically. For example, adult learners are incapable of perceiving the difference and forming new phonetic representations in memory for similar, but not identical, L2 sounds and substitute them with L1 sounds. This was supported by observation of experienced L2 learners who lived in the target language environment for about five years and preserved a strong interference of L1. These observations supported the view that only in supervised learning can adult learners develop accurate pronunciation and not merely through extensive interaction experience without sufficient feedback.

The model of learning sound production reflected the principle of close dependence between how sounds were stored and how they are produced. Sound production is a separate modality in memory and requires an individual input different from that required for sound recognition learning. Nevertheless, strongly accented sound production may affect the way the sounds are stored in memory. For example, if a student learns to recognize sounds accurately by receiving a sufficient authentic auditory input, supported by theoretical explanations of the differences between L1 and L2 sounds, he/she can still develop inaccuracy in sound recognition due to the input coming from his/her inaccurate production. This suggests that developing accurate articulation movements is as important as the authentic storing of phonological information. Developing recognition and articulation of new sounds of Mandarin phonology requires an instructional program to provide prolonged training designed for developing new articulation behavior.

Based on the models of memory for learning L2 pronunciation, we analyzed the teaching Mandarin Chinese pronunciation in the Department of East Asian Languages and Literatures. An analysis of the most common mispronunciations on different levels of learning, as reported by instructors with extensive experience in teaching the Chinese language, allowed us to propose guiding principles for teaching Chinese pronunciation. As a framework for the teaching Chinese pronunciation program, we took the Intensive Chinese Curriculum Design created by Galal Walker (1989) and proposed some guidelines for including instruction in pronunciation throughout an entire program of teaching Chinese. The guidelines and the exemplary exercises rest upon the basic knowledge of memory and mnemonic techniques discussed in Chapter 2 and the proposed models of learning L2 pronunciation.

List of References

A Practical Course in Chinese Phonetics. 1982. Beijing yuyan xueyuan.

Anderson, J.R., and G.H. Bower. 1972. Recognition and retrieval processes in free recall. *Psychological Review*, 79: 97-123.

Anisfeld, M., and M.E. Knapp. 1968. Association, synonymity and directionality in false recognition. *Journal of Experimental Psychology*, 77: 171-179.

Atkinson, R.C., and R.M. Shiffrin. 1968. Human Memory: A proposed system and its control processes. In K. W. Spence and J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in Research and Theory.* Vol. 2. New York: Academic Press.

Atkinson, R.C., and R.M. Shiffrin. 1971.The control of short term memory. *Scientific American*, 225: 82-90.

Baddeley A.D., and B.Wilson. 1985. Phonological coding and short-term memory in patients without speech. *Journal of Memory and Language*, 24: 490-502.

Baddeley, A. 1966. The influence of acoustic and semantic similarity in long-term memory for word sequences. *Quarterly Journal of Experimental Psychology*, 18: 362-365.

Baddeley, A.D. 1976.*The Psychology of Memory.* Basic Books, Inc., Publishers New York. Ed. S. Glucksberg.

Baddeley, A.D. 1970. Effects of of acoustic and semantic similarity on short-term pair-associate learning. *Journal of Psychology*, 61, 335-343.

Baddeley, A.E. Working memory. Oxford, England: Clarendon Press. 1986.

Barnard, P. 1985. Interactive cognitive subsystem: A psycholinguistic approach to short-term memory. In A.W. (Ed.), *Progress in the Psychology of Language*. Vol.2 197-258, London: Lawrence Erlbaum Press.

Bechtel, W., and A. Abrahamsen. 1991. *Connectionism and the Mind. An Introduction to Parallel Processing in Networks.* Basil Blackwell, Inc. 3 Cambridge Center, Cambridge, Massachusets.

Bjork, and , A.F. Healy. 1974. Short-term order and item retention. *Journal of Verbal Learning and Verbal Behavior*, 13: 80-97.

Bousfield, W.A. 1953. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 49: 229-240.

Bousfield, W.A., and B.M. Cohen.1955.The occurrence of clustering in the recall of randomly arranged words of different frequencies of usage. *Journal of Genetic Psychology*, 52: 83-95.

Bower, G.H. and Clark, M.C. 1969. Narrative stories as mediators for serial learning. *Psychonomic Science*, 14: 181-182.

Bower, G.H. and D. Winzenz . 1969. Group structures, coding and memory for digit series. *Journal of Experimental Psychology Monograph*, 80, No. 2, Part 2: 1-17.

Bower, G.H. Mental imagery and associative learning . 1973. In L. Gregg (Ed.), *Cognition in Learning and Memory*. New York: John Wiley and Sons.

Bower, G.H. 1970. Organizational factors in memory. *Cognitive Psychology*, 1:18-46.

Briggs, G. 1954. Acquisition, extinction and recovery functions in retroactive inhibition. *Journal of Experimental Psychology*, 47: 285-293.

Brown, J. 1958.Some tests of decay theory of immediate memory. *Quarterly Journal of Experimental Psychology*, 10: 12-21.

Buschke, H. and R. Lenon. 1969.Encoding homophones and synonyms for verbal discrimination and recognition. *Psychonomic Science*, 14: 269-270.

Conrad, R. Acoustic confusions in immediate memory. 1964. *British Journal of Psychology*, 55: 75-84.

Conrad, R. and A. J. Hull. 1964. Information , acoustic confusion and memory span. *British Journal of Psychology*, 55: 429-432.

Conrad, R. 1967. Interference or decay over short-term retention intervals? *Journal of Verbal Learning and Verbal Behavior*, 6: 49-54.

Craik, F.I.M., and R.S. Lockhart. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 11: 671-684.

Craik, F.I.M., and M. J. Watkins. 1973. The role of rehearsal in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 12: 598-607.

Darwin, C.J., M.T. Turvey and R.G. Crowder. 1972. An auditory analogue of the Sperling partial-report procedure: Evidence for brief auditory storage. *Cognitive Psychology*, 3: 255-267.

Delattre, P. 1964.Comparing the vocalic features of English, German, Spanish and French. *International Review of Applied Linguistics*. 7: 295-323.

Delattre, P. 1969. An acoustic and articulatory study of vowel reduction in four languages. *International Review of Applied Linguistics*. 7: 295-323.

Donahoe, J. W. and M.G. Wessels. 1980. *Learning, Language and Memory*. Harper and Row Publishers. New York.

Flege, J.E. 1988. The development of skill in producing word-final /p/ and /b/: Kinematic parameters. *Journal of Acoustical Society of America*, 84: 1639-1652.

Flege, J.E. 1989. Chinese subject's perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of Acoustical Society of America*, 2: 107-128.

Flege, J.E. 1991. Perception and Production: The relevance of phonetic input to L2 Phonological Learning. In T. Heubner and C. Ferguson (Eds.) *Crosscurrents in Second Language Acquisition and Linguistic Theory*. Philadelphia: John Benjamins.

Flege, J.E. 1992. Speech learning in a second language. In C. Ferguson, L. Menn and C. Stoel Gammon (Eds.) *Phonological Development: Models, Research and Applications*, Parkton. M.D: York Press, 565-604.

Flege, J.E., Munro, M., and Fox, R. 1992. The interlingual identification of Spanish and English vowels. *Journal of the Acoustical Society of America*.

Flege, J.E. and Wang, C. Native language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/- /d/ contrast. Journal of Phonetics, 17, 292-315, 1990.

Flege, J.E., and J. Hillenbrand 1984. Limits on phonetic accuracy in adult foreign language speech production. *Journal of the Acoustical Society of America* 79:708-721.

Flege, J.E., and J. Hillenbrand. 1987. Differential use of closure voicing and release burst as cue to stop voicing by native speakers of French and English. *Journal of Phonetics* 15, 203-208.

Flege, J.E., and J. Hillenbrand. 1987. Limits on phonetic accuracy in foreign language speech production. In G. Ioup and S. H. Weinberger (Eds.) *Interlanguage Phonology: The Acquisition of Second Language Sound System*. 176-203. Newbury House Publishers.

Genessee, F. 1988.Neuropsychology and second language acquisition. In Leslie M. Beebe(ed.) *Issues in Second Language Acquisition. Multiple perspectives*. Columbia University, 81-112.

Grossman, L. and M. Eagle. 1970. Synonymity, antonymity and association in false recognition responses. *Journal of Experimental Psychology,* 83, 244-248.

Hammerly, H. 1991.*Fluency and Accuracy: Toward balance in Second Language Teaching and Learning*. Multilingual matters Ltd. Clevedon, Philadelphia, Adelaide.

Hayes, J.R.M. 1952. Memory span for several vocabularies as a function of a vocabulary size. In *Quarterly Progress Report*. Cambridge, Massachusets: Acoustic Laboratory, Massachusets Institute of Technology, Jan-Jun.

Hebb, D.O. 1961. Distinctive features of learning in the higher animal. In J. Delafresnaye (Ed.) *Brain Mechanisms and Learning*. London and New York: Oxford University Press.

Hecht, B. F., R . Mulford. 1987. The acquisition of a second language phonology: Interaction of transfer and developmental factors In G. Ioup and S. H. Weinberger (Eds.) *Interlanguage Phonology: The Acquisition of Second Language Sound System,* 213-229. Newbury House publishers.

Hellyer, S. 1962. Frequency of stimulus presentation and short-term decrement in recall. *Journal of Experimental psychology*, 64: 650.

Hintzman, D.L. 1967.Articulatory coding in short term memory. *Journal of Verbal Learning and Verbal Behavior*, 6: 312-316.

Hintzman, D. 1983. *Schema Abstraction in a Multiple Trace Memory Model*. Paper presented at a conference on "The prioroty of specific", Elora, Ontario, Canada.

Howard, D. and S.Franklin.1987. Implications for lexical theories. *In Missing the meaning? A Cognitive Neuropsychological Study of the Processing of Words by an Aphasic Patient*. A Bradford Book. The MIT Press. Cambridge, Massachusets; London, England, 111-128.

Jacoby, L.L. 1983a. Perceptual enhancement: Persistent effects of an experience. *Journal of an Experimental Psychology: Learning, Memory and Cognition*, 9: 21-38.

Jacoby, L.L. 1983b. Remembering the data: Analyzing interaction processes in reading. *Journal of Verbal Learning and Verbal Behavior*, 22: 485-508.

Kintsch, W. 1968.An experimental analysis of single stimulus tests and multiple-choice tests of recognition memory. *Journal of experimental psychology,* 76: 1-6.

Kintsch, W. 1970 (a) *Learning, Memory and Conceptual Processes*. New York: Wiley.

Kintsch, W. 1970. (b). Models for free recall and recognition . In D. A. Norman (Ed.), *Models of Human Memory*. New York: Academic Press.

Kintsch, W., and H. Buschke 1969. Homophones and synonyms in short-term memory. *Journal of Experimental Psychology*, 80, 403-407.

Klapp, S.T., E. A. Mashburn and P. T. Lester. 1983. Short term memory does not involve the "working memory" of information processing. The demise of a common assumption. *Journal of Experimental psychology*. General. 112: 240-241.

Krashen, S.D. 1974. The critical period for language acquisition and its possible bases. *Annals of the New York Academy of Sciences*, 263: 211-224.

Kratochvil, Paul. 1968. *The Chinese laguage today. Features of an emerging standard.* Hutchinson university library. London.

Krause, S.E. 1982a. Developmental uses of vowel duration as a cue to post-vocalic stop voicing. *Journal of Speech and Hearing Research,* 25: 388-93.

Krause, S.E. 1982b.Vowel duration as a perceptual cue to post-vocalic voicing in young children and adults. *Journal of Acoustical Society of America*. 71: 990-995.

Lamendella, J. 1977. General principles of neuro-functional organization and their manifestation in primary and and non-primary language acquisition. *Language Learning,* 27: 155-196.

Lenneberg, E. *Biological Foundations of Language*. New York, Wiley, 1967.

Lindsay, P.H. and D.A. Norman. 1972. *Human Information Processing: An Introduction to Psychology*. New York: Academic Press.

Loftus G.R. and E.F. Loftus.1976. *Human Memory. The Processing of Information.* Lawrence Erlbaum Associates, Publishers Hillsdale, New Jersea.

Lukatela, G. and Turvey, M.T. 1994. Visual lexical access is initially phonological:1. Evidence from associative Priming by words, homophones and pseudohomophones. *Journal of Experimental Psychology: General*. Vol. 123, No 2: 107-128.

Martin, R.C., J.R. Shelton and L.S. Yaffee.1994. Language processing and working memory. Neuropsychological evidence for separate phonological and semantic capacities. *Journal of Memory and Language*, 33: 83-111.

McCelland J.L. and Rumelhart D.E. 1986. A Distributed Model of Human Learning and Memory. In: McClelland J.L. and Rumelhart, D.E., & the PDP Research Group. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition.* Vol. 2. Psychological and Biological Models. Cambridge, MA: MIT Press/Bradford Books, 170-215.

Miller, G.A. 1956. The magical number seven, plus or minus two: Some limits on our capacity to process information. *Psychological Review*, 9: 81-97.

Molfese, D.L. and V.J. Molfese.1979. Hemisphere and stimulus differences as reflected in the cortical responces of newborn infants to speech stimuli. *Developmental Psychology*, 15: 505-511.

Mousell, S. 1984.Components of working memory underlying verbal skills: A "distributed capacities" view. In: *International Symposium on Attention and Performance.* edited by H. Bouma and D. Boumhuis. Hillsdale, N. J. Erlbaum, 327-350.

Nelson, T.O. and Rothbart, R. 1972. Acoustic savings for items forgotten from long-term memory. *Journal of Experimental Psychology*, 93: 357-360.

Neufeld, G. and E. Schneidermann. 1980. Prosodic and articulatory features in adult language learning. In R.C. Scarceland and S. D. Krashen (Eds.), *Research in Second Language Acquisition.* Rowley, MA: Newbury House.

Norman, Jerry. 1988. *Chinese.* (Cambridge University Surveys). Cambridge University Press.

Moray, N., Bates, A., and Barnett. 1965. I. Experiments on the four-eared man. *Journal of the Acoustical Society of America*, 38: 196-201.

Munro, M. 1965. Attention to spectral and temporal cues in vowel perception among native speakers of Arabic and English. *Journal of the Acoustical Society of America,* 88: S53(A).

Parallel Models of Asssociative Memory. 1981. Eds.G.E. Hinton and S.A. Anderson.

Pavlov, I.. 1927. *Conditioned Reflexes*. London and New York: Oxford University Press.

Penfield, W., and Roberts, L. 1959. *Speech and brain mechanisms*.New York. Atheneum.

Pollack, I. 1953. The assimilation of sequentially coded information. *American Journal of Psychology*, 66, 421-435.

Postman, L., P.A. Adams and L.W. Philips. 1955. Studies in incidental learning II. The effects of association value and method of testing. *Journal of Experimental Psychology,* 49, 1-10.

Postman, L., 1950. Choice behavior and the process of recognition. *American Journal of Psychology*, 63: 443-447.

Rosch, E. 1975. Cognitive representations of semantic behavior. *Journal of Experimental Psychology: General*, 104: 192-233.

Rundus, D. 1971. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 89: 63-77.

Rundus, D. and R.C.Atkinson. 1970. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 9: 99-105.

Scovel, T. 1988. *A time to speak.* A psycholinguistic inquiry into the critical period for human speech. New York, Newbury House,.

Seliger, H.W. 1978. Implications of a multiple critical periods hypothesis for second language learning. In W.C. Ritchie (Ed.) *Second Language Acquisition Research.* New York Academy Press.

Shiffrin, R. M. 1970 (a). Forgetting: Trace erosion or retrieval failure? *Science,* 168, 1601-1603.

Shiffrin, R. M. Memory Search. 1970 (b). In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press.

Shulman, H.G. 1972. Seminar confusion errors in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 11: 221-227.

Simpson, G.B.and M.A. Krueger.1991. Selective access of homograph meanings in sentence context. *Journal of Memory and Language*, 30: 627-643.

Sperling, G. A. 1963.A model for visual memory tasks. *Human Factors*, 5, 19-39.

Sperling, G.A. 1967. Successive approximations to a model for short-term memory. *Acta Psychologica*, 27, 285-292.

Sperling, G. and R.G. Speelman. 1970. Acoustic similarity and auditory short-term memory: Experiments and a model. In D. A. Norman (Ed.), *Models of Human Memory.* New York: Academic Press.

Sperling, G., J.Budiansky, L. Spivak, and M.C Johnson.1971.The maximum rate for scanning letters for the presence of a numeral. *Science,* 174: 307-311.

Squire, L.R. 1987. *Memory and Brain.* Oxford University Press. New York. Oxford.

Sternberg, S. 1969.Mental processes revealed by reaction-time experiments. *American Scientist,* 57: 421-457.

Sternberg, S. 1966. High-speed scanning in human memory. *Science,* 153: 652-654.

Sternberg, S. 1967. Two operations in character recognition: Some evidence from reaction-time measurements. *Perception and Psychophysics*, 2: 45-53.

Stevick, E.W. 1976. *Memory, Meaning and Method.* Some psychological perspectives on language learning. Newbury House Publishers, Inc., Rowley, Massachusets.

Tulving, E. 1972. Episodic and semantic memory. In E. Tulving and W. Donaldson (eds.) *Organization of Memory.* New York: Academic Press.

Tulving, E. 1962. Subjective organization in free recall of "unrelated words". *Psychological Review.* 69, 344-354.

Tulving, E. and Z. Pearlstone. 1973. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior.* 5: 381-391.

Tulving, E., and D.M. Thompson 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological Review,* 80: 352-372.

Tulving, E., and D. M. Thompson. 1971. Retrieval Processes in recognition memory; Effect of associative context. *Journal of Experimental Psychology,* 87: 116-124.

Underwood, B.J. 1964. Forgetting. *Scientific American,* reprint 482.

Vygotsky, L. 1962. *Thought and Language.* Cambridge, MA: MIT Press.

Walker, G. 1982. Videotext: A course in intermediate to advanced Chinese. *Journal of Chinese Language Teachers Association,* 17.2:109-122.

Walker, G. 1987. "Literacy" and "Reading" in a Chinese Pogram. *Journal of Chinese Language Teachers Association,* 19.1: 67-84.

Walker, G. Intensive Chinese curriculum: The EASLI model. *Journal of the Chinese Language Teachers Association,* 24:2,43-83, 1989.

Walker, G.L.R., Zhang Zhanyi, Huang Zhengcheng and Lu Jianji.1987. *A self-study Introduction to Beginning Mandarin Chinese,* Vol.1-4. Chinese Individualized Instruction Project. Department of East Asian Languages and Literatures. The Ohio State University.

Walsh, T.M. and K.C. Diller. 1981. Neurolinguistic considerations on the optimum age for second language learning. In K. Diller (Ed), *Individual Differences and Universals in Language Learning Aptitude.* Rowley, MA: Newbury House.

Wenk, B. J. Speech rhythms.1985. *Language and Speech,* 28:157-175.

Whittlesea, B.W.A. 1983. *Representation and Generalization of Concepts: The Abstractive and Episodic Perspectives Evaluated.* Unpublished Doctoral Dissertation, MacMaster University.

Wickens, D.D. 1972. Characteristics of word encoding. In A.W. Melton and E. Martin (Eds.), *Coding Processes in Human Memory.* Washington D.C.:Winston.

Watkins, M.J., and O.C Watkins.1974. Processing of recency items for free recall. *Journal of Experimental Psychology,* 102: 488-493.

Wode, H. 1978. The beginning of non-school room L2 phonology acquisition. In *International Review of Applied Linguistics,* 16: 109-125.

Wode, H. *Learning a Second Language. 1. An integrated View of Language Acquisition.* Gunter Narr Verlag. Tubingen 1981.

Woodward, A.E., R.A. Bjork and R. H.Longeward. 1973 Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior*, 12, 608-617.

Yi Xu. Depth of phonological recoding in short-term memory. 1991. *Memory & Cognition,* 19(3): 263-273.

Zhang Guojun and H.A. Simon. 1985. STM capacity for Chinese words and idioms: Chunking and acoustical loop hypotheses. *Memory and Cognition,*13 (3), 193-201.